

I. Zwiener¹, M. Blettner¹, G. Hommel¹

Überlebenszeitanalyse – Teil 15 der Serie zur Bewertung wissenschaftlicher Publikationen

*Survival analysis – part 15 of a series on evaluation
of scientific publications*

Hintergrund: Zum Vergleich von zwei Therapien werden häufig Überlebenszeiten herangezogen. Diese sollten korrekt ausgewertet und interpretiert werden können.

Methoden: Anhand einer publizierten Studie zur Behandlung von Patienten mit Hirntumoren werden spezielle Verfahren zur Auswertung von Überlebenszeitdaten beschrieben. Diese Arbeit basiert auf ausgewählten Lehrbüchern der Statistik, einer selektiven Literaturlauswahl und der eigenen Expertise.

Ergebnisse: Bei der Analyse von Überlebenszeitdaten wird das Kaplan-Meier-Verfahren benutzt. Es können Überlebensraten und die mediane Überlebenszeit angegeben werden. Mit Hilfe des Log-rank-Tests kann man die Überlebenszeiten von zwei Gruppen miteinander vergleichen. Für multivariable Modelle verwendet man die Cox-Regression. Das Hazard Ratio als deskriptives Maß für den Unterschied von Überlebenszeiten wird erläutert.

Schlussfolgerungen: Wenn nicht spezielle Verfahren bei der Analyse von Überlebenszeitdaten eingesetzt oder deren Annahmen nicht überprüft werden, können die Ergebnisse fehlerhaft sein. Der Leser einer wissenschaftlichen Publikation sollte diese Fallstricke kennen und beurteilen können, ob die gewählte Auswertestrategie korrekt ist.

(Dtsch Zahnärztl Z 2011, 66: 762–769)

Background: Survival times are often used to compare treatments. Survival data are a special type of data, and therefore have to be analyzed with special methods.

Methods: We illustrate special techniques for analyzing survival times by applying them to a publication on the treatment of patients with brain tumors. The present article is based on textbooks of statistics, a selective review of the literature, and the authors' own experience.

Results: Survival times are analyzed with the Kaplan-Meier method, which yields two measures of interest: survival rates and the median survival time. The log-rank test is used to compare survival times across treatment groups. Cox regression is used in multivariable models. The hazard ratio, a descriptive measure for differences in survival times, is explained.

Conclusion: If survival times are analyzed without the use of special techniques, or if the underlying assumptions are not taken into account, faulty interpretation may result. Readers of scientific publications should know these pitfalls and be able to judge for themselves whether the chosen analytical method is correct.

* Nachdruck aus: Dtsch Arztebl Int 2011; 108(10): 163–169; DOI: 10.3238/arztebl.2011.0163 © Deutscher Ärzte-Verlag GmbH Köln

¹ Institut für Medizinische Biometrie, Epidemiologie und Informatik (IMBEI), Universitätsklinikum Mainz

Peer reviewed article: eingereicht: 1.6.2010, revidierte Fassung angenommen: 12.10.2010

DOI: 10.3238/dzz.2011.0762

Kasten 1: Typische Fehler bei der Überlebenszeitanalyse**1. Auswerten der rohen Ereignishäufigkeiten**

- Für jeden Patienten wird nur betrachtet, ob während der Studienlaufzeit ein Ereignis beobachtet wurde oder nicht. Es wird nicht betrachtet, wann dieses Ereignis aufgetreten ist, oder wie lange die Patienten ohne Ereignis beobachtet wurden.
→ Problem: Beim Therapievergleich zählen nur die Häufigkeiten der beobachteten Ereignisse. Dies ist im Allgemeinen falsch, da die Zeit bis zum Ereignis außer Acht gelassen wird. Bei Patienten mit langer Nachbeobachtungszeit wird man häufiger ein Ereignis sehen können als bei Patienten mit kurzer Nachbeobachtungszeit.
- Beispiel: Die letzten 10 Jahre wurde zur Hirntumorbehandlung eine Standardtherapie eingesetzt. Seit einem Jahr wird eine neue Therapie angewendet (maximale Beobachtungszeit der Patienten mit neuer Behandlung: 1 Jahr). Nun möchte man untersuchen, ob unter der neuen Therapie weniger Patienten sterben. Da viele der Patienten auch noch 2 oder 3 Jahre nach Diagnose ihr Leben verlieren, werden künftig noch weitere Patienten, die die neue Behandlung bekommen, sterben. Die Betrachtung der rohen Ereignishäufigkeiten würde zu einem verzerrten Ergebnis führen.

2. Ausschluss der zensierten Patienten

- Nur Patienten, bei denen ein Ereignis eingetreten ist, gehen in die Auswertung ein, zensierte Patienten (ohne Ereignis) werden aus der Analyse ausgeschlossen. Die Zeit bis zum Ereignis wird mittels eines t-Tests verglichen.
→ Problem: Zensierte Patienten sind Patienten, bei denen während der gesamten Beobachtungszeit kein Ereignis eingetreten ist. Dies ist eine wichtige Information, die nicht aus der Analyse ausgeschlossen werden darf.
- Beispiel: Unter Therapie A sterben innerhalb eines Jahres 10 % der Patienten, unter Therapie B kommen innerhalb eines Jahres 50 % der Patienten zu Tode. Wenn man nur betrachtet, wann die 10 % bzw. 50 % der Patienten aus dem Leben traten, sind beide Therapien gleich gut (alle innerhalb des ersten Jahres verstorben). Die Information, wie viele Patienten nicht verstorben sind (also zensiert), muss aber auch berücksichtigt werden.

3. Zensierungszeit = Ereigniszeit

- Es gehen alle Patienten in die Auswertung ein. Es wird aber nicht unterschieden, ob bei einem Patienten ein Ereignis auftrat oder ob der Patient zensiert ist. Die Überlebenszeit (zensiert oder mit Ereignis) wird mit einem t-Test verglichen.
→ Problem: Die Ergebnisse sind verzerrt, da Zensierungen als Ereignisse ausgelegt werden.
- Beispiel: Aufgenommen wird wieder das Beispiel unter Punkt 1. Unter der Standardbehandlung gibt es Patientendaten über eine Zeit von bis zu 10 Jahren. Für die neue Therapie liegen nur Daten über 1 Jahr vor. Wenn man nun alle zensierten Patienten als Patienten, bei denen ein Ereignis eingetreten ist, auslegt, sieht es so aus, als ob die neue Therapie schlechter wäre, weil alle Überlebenszeiten maximal 1 Jahr lang sind (unter der Standardbehandlung gibt es Patienten mit bis zu 10 Jahren Überlebenszeit).

Einleitung

In vielen Bereichen der Medizin ist die primäre Zielgröße die Zeit bis zum Auftreten eines Ereignisses. Sie kann beispielsweise die Zeit von Lungenkrebsdiagnose bis zum Tod, die Zeit vom Einsetzen einer Zahnprothese bis zur ersten Reparatur oder die Zeit von Beginn einer Harninkontinenzbehandlung bis zum Therapieerfolg umfassen. Das „Ereignis“ kann sowohl ein Erfolg (Heilung) als auch ein Misserfolg (Tod) sein. Wichtig ist, dass sowohl der Startzeitpunkt als auch der Ereigniszeitpunkt klar definiert sind. Die Zeitspanne dazwischen wird allgemein als Überlebenszeit bezeichnet, selbst wenn das eintretende Ereignis nicht der Tod ist.

In fast allen medizinischen Fachzeitschriften findet man Artikel, in denen

Techniken der Überlebenszeitanalyse benutzt werden. Ein erst kürzlich erschienenes Beispiel ist eine Studie über Patienten mit einem Hirntumor. *Von Hoff et al.* [12] untersuchten 280 Kinder und Jugendliche mit einem Medulloblastom aus der zweiarmligen, randomisierten HIT'91-Studie (HIT, Hirntumor). Patienten in Arm 1 erhielten eine Chemotherapie vor und nach der Bestrahlung (Sandwich-Chemotherapie), Patienten aus Arm 2 bekamen sofort eine Bestrahlung, gefolgt von einer Chemotherapie (Erhaltungschemotherapie). Es wurde überprüft, ob eine der zwei Therapien zu längeren Überlebenszeiten der Patienten führt.

Um die Ergebnisse und die Wertigkeit solcher Publikationen richtig interpretieren zu können, sollte der Leser mit

den Verfahren der Überlebenszeitanalyse vertraut sein. Die vorliegende Arbeit führt anhand der HIT'91-Studie schrittweise in die Techniken der Überlebenszeitanalyse ein und versetzt den Leser in die Lage, diese selbst verstehen und interpretieren zu können.

Besonderheit von Überlebenszeitdaten

Klinische Studien sind sowohl aus ethischer als auch aus finanzieller Sicht zeitlich begrenzt. Das erwartete Ereignis, zum Beispiel Tod oder Therapieerfolg, tritt bei einigen Patienten erst nach Studienende oder auch gar nicht ein. Daher liegt für diese lediglich die Information vor, dass bis zu einem be-

Kasten 2: Kaplan-Meier-Verfahren. Beispiel anhand der Daten von 5 Kindern mit einem Hirntumor

Die Patientendaten sind in der Tabelle 1 nach der Beobachtungszeit geordnet. Drei der fünf Patienten starben.

Pat-Nr.	Tod	Überlebenszeit t_i in Monaten	n_i	d_i	Kaplan-Meier-Schätzer $S(t_i)$
3	Ja	4	5	1	$4/5 = 80\%$
5	Ja	7	4	1	$4/5 \times 3/4 = 60\%$
2	Nein	9	3	0	
1	Ja	10	2	1	$4/5 \times 3/4 \times 1/2 = 30\%$
4	Nein	13	1	0	

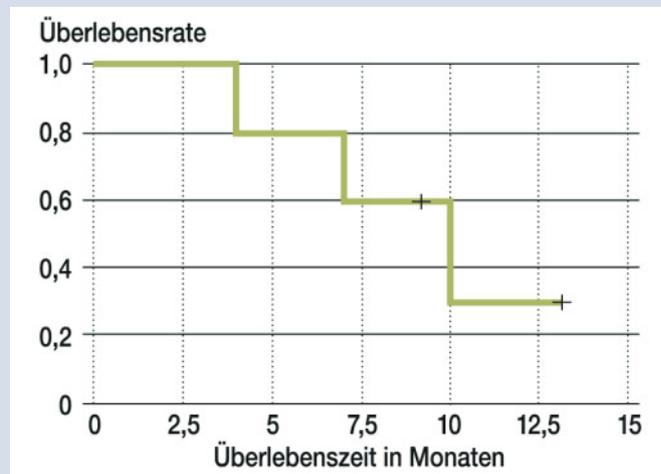
Tabelle 1 Arrangement der Überlebenszeiten und Kaplan-Meier-Schätzer. t_i : i-ter Ereigniszeitpunkt; n_i : Anzahl der Patienten in der Risikomenge zum Zeitpunkt t_i ; d_i : Anzahl der Patienten, bei denen zum Zeitpunkt t_i ein Ereignis eingetreten ist; $S(t_i)$: Kaplan-Meier-Schätzer zum Zeitpunkt t_i .

Table 1 Survival times and Kaplan-Meier estimate.

t_i : time of event no. i ; n_i : no. of patients at risk at time t_i ; d_i : no. of patients who have suffered an event by time t_i ; $S(t_i)$: Kaplan-Meier estimate of the survival function at time t_i .

Berechnung des Kaplan-Meier-Schätzers (Tabelle 1)

- 4. Monat: 1 von 5 Patienten stirbt → Wahrscheinlichkeit mindestens bis zum Ende des vierten Monats zu überleben = $4/5 = 80\%$
- 7. Monat: 1 von 4 Patienten stirbt (vom 4. bis zum 7. Monat überleben $3/4$) → Wahrscheinlichkeit mindestens bis zum Ende des siebten Monats zu überleben = $4/5 \times 3/4 = 60\%$ (Wahrscheinlichkeit insgesamt mindestens bis zum Ende des siebten Monats zu überleben ist das Produkt aus den beiden vorhergehenden Wahrscheinlichkeiten)
- 9. Monat: ein zensierter Patient (während der Studienlaufzeit trat bei dem Patient kein Ereignis auf) → Risikomenge verringert sich auf 2 Patienten, Kaplan-Meier-Schätzer bleibt aber auf konstantem Niveau (da zu diesem Zeitpunkt niemand verstorben ist)
- 10. Monat: 1 von 2 Patienten stirbt → Wahrscheinlichkeit mindestens bis zum Ende des zehnten Monats zu überleben = $4/5 \times 3/4 \times 1/2 = 30\%$
- 13. Monat: ein zensierter Patient → keine weiteren Patienten in der Risikomenge → Kaplan-Meier-Schätzer endet und damit auch die Kaplan-Meier-Kurve



Grafik 1 Kaplan-Meier-Kurve.

Figure 1 Kaplan-Meier curve.

Kaplan-Meier-Kurve (Grafik 1)

Zu jedem Todeszeitpunkt macht die Kaplan-Meier-Kurve einen Sprung nach unten, zensierte Patienten sind durch einen senkrechten Strich gekennzeichnet. Beim Kaplan-Meier-Verfahren werden zensierte Patienten berücksichtigt:

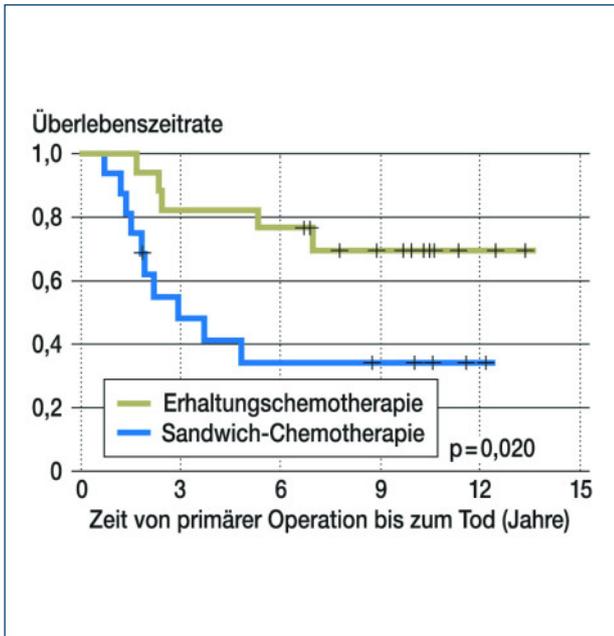
- Wenn man annimmt, dass Patient Nr. 2 verstorben wäre, hätte nur 1 von 5 Patienten die maximale Beobachtungszeit von 13 Monaten überlebt, also 20 %.
- Wenn man annimmt, Patient Nr. 2 hätte weiterhin überlebt, hätten 2 von 5 Patienten die Beobachtungszeit von 13 Monaten erreicht, also 40 %.
- Es ist aber unsicher, was mit dem zensierten Patient weiterhin geschah. Der Kaplan-Meier-Schätzer spiegelt dies durch die Schätzung der Überlebensrate von 30 % wider.

stimmten Zeitpunkt noch kein Ereignis eingetreten ist. Diese Art von Beobachtung wird als Zensierung bezeichnet. Eine Zensierung kann auch auftreten, wenn Personen die Studie verlassen. Dies passiert zum Beispiel, wenn sie nicht weiter an der Studie teilnehmen möchten oder wenn sie aus Gründen,

die nicht mit der Studie im Zusammenhang stehen, sterben.

In der Onkologie wird oft zwischen dem Gesamtüberleben (Zeit von der Diagnose bis zum Tod jedweder Art) und dem tumorspezifischen Überleben (Zeit von Diagnose bis zum Tod infolge des Tumors) unterschieden. Beim tumorspe-

zifischen Überleben sind Patienten, die nicht aufgrund des Tumors sterben, zensiert, weil das Ereignis „Tod durch Tumor“ nicht eingetreten ist. In komplexeren Auswertungen können auch beide Ereignisse parallel untersucht werden (als konkurrierende Risiken). Darauf soll aber in dieser Arbeit nicht eingegangen



Grafik 2 Kaplan-Meier-Kurve für 33 Kinder und Jugendliche mit einem Medulloblastom und Metastasenstatus M1.

Erhaltungstherapie: 10-Jahres-Überlebensrate = 70 %, mediane Überlebenszeit nicht bestimmbar; Sandwich-Chemotherapie: 10-Jahres-Überlebensrate = 36 %, mediane Überlebenszeit = 2,9 Jahre (Aus: von Hoff K et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT'91. EJC 2009; 45: 1209-17 [12]; Abdruck mit freundlicher Genehmigung des Verlags Elsevier, Oxford).

Figure 2 Kaplan-Meier curve for 33 children and adolescents with medulloblastoma and metastasis status M1.

Maintenance chemotherapy: 10-year survival rate = 70 %, median survival rate cannot be determined Sandwich chemotherapy: 10-year survival rate = 36 %, median survival rate = 2.9 years (From: von Hoff K et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT'91. EJC 2009; 45: 1209-17 [12]; printed with the kind consent of Elsevier Publishers, Oxford).

werden. In der HIT'91-Studie wird die Zeit von primärer Operation des Hirntumors bis zum Tod jedweder Art betrachtet.

Neben den Daten von Patienten mit bekannter Überlebenszeit müssen auch die von zensierten Patienten in die Auswertung eingehen. Um die Daten der zensierten Patienten adäquat in die Analyse einfließen zu lassen, sind besondere Auswertungsstrategien notwendig.

Werden Überlebenszeitdaten nicht mit diesen Verfahren ausgewertet, sind die Ergebnisse im Allgemeinen fehlerbehaftet. Die häufigsten Fehler bei der Auswertung von Überlebenszeitdaten werden in Kasten 1 vorgestellt.

Bei der Auswertung von Überlebenszeiten ist es wichtig, sowohl die Zeit bis zum Auftreten des Ereignisses als auch die Zensierungen zu berücksichtigen. Verfahren zur Auswertung und grafischen Darstellung von Überlebenszeitdaten werden im Folgenden anhand der HIT'91-Studie vorgestellt. Einfache Einführungen in die Überlebenszeitanalyse geben die Lehrbücher von Weiß [13] und von Schumacher und Schulgen [11]. Zur weiterführenden Lektüre sei auf die Lehrbücher von Collett [2] oder Kalbfleisch und Prentice [5] verwiesen.

Kaplan-Meier-Kurve

Die Tabelle 1 in Kasten 2 stellt die Überlebenszeiten von fünf Kindern mit ei-

nem Hirntumor dar. Die Wahrscheinlichkeit, dass ein Patient bis zu einem bestimmten Zeitpunkt überlebt hat, wird durch das Kaplan-Meier-Verfahren berechnet [6]. Grafisch können die Überlebenszeiten mittels einer Kaplan-Meier-Kurve (auch Überlebenszeitkurve genannt) dargestellt werden (Grafik 1 in Kasten 2). Dazu trägt man auf der x-Achse die Überlebenszeit der Patienten und auf der y-Achse die mit dem Kaplan-Meier-Verfahren berechnete Überlebenswahrscheinlichkeit ab.

Die Berechnung der Überlebenswahrscheinlichkeiten und die grafische Darstellung durch die Kaplan-Meier-Kurve werden in Kasten 2 Schritt für Schritt erläutert.

Überlebensraten und mediane Überlebenszeit

Mit Hilfe der Kaplan-Meier-Kurve lassen sich Überlebensraten bestimmen. Überlebensraten geben an, bei wie vielen Patienten bis zu einem bestimmten Zeitpunkt noch kein Ereignis eingetreten ist. Im obigen Beispiel beträgt die 1-Jahres-Überlebensrate 30 % (Kasten 2). Dies lässt sich folgendermaßen interpretieren: Ein Jahr nach Diagnose erwartet man, dass noch 30 % der Patienten leben. Bei der Angabe von Überlebensraten ist es wichtig, den zugehörigen Zeitpunkt mit anzugeben. Zum Vergleich von zwei Therapiegrup-

pen wird empfohlen die Kaplan-Meier-Kurven beider Therapiegruppen darzustellen, da diese mehr Information enthalten als die Angabe einzelner Überlebensraten.

Die mittlere Überlebenszeit hängt sehr stark vom Zensierungsmuster ab, so dass bei Überlebenszeiten immer der mediane Wert angegeben wird. Die mediane Überlebenszeit ist der Zeitpunkt, zu dem die Hälfte der Patienten ein Ereignis erlitten hat. Bei den fünf Hirntumorpatienten beträgt die mediane Überlebenszeit zehn Monate. Falls der Kaplan-Meier-Schätzer in der gesamten Beobachtungszeit über 50 % liegt, ist die mediane Überlebenszeit nicht zu bestimmen. In diesem Fall ist bis zur maximalen Beobachtungszeit für weniger als die Hälfte der Patienten ein Ereignis eingetreten.

Log-rank-Test

In der HIT'91-Studie werden die Überlebenszeiten der Patienten aus den zwei Therapiegruppen getrennt für die einzelnen Metastasenstadien verglichen. Zum deskriptiven Vergleich der Überlebenszeiten der beiden Therapiegruppen innerhalb der Patienten mit Metastasenstatus M1 dienen die Kaplan-Meier-Kurven (Grafik 2). Um die Überlebenszeiten statistisch miteinander zu vergleichen, wird das Standardverfahren, der Log-rank-Test, ange-

Kasten 3: Hazard and Hazard Ratio**Hazard $h(t)$**

Das Risiko, genau zum Zeitpunkt t ein Ereignis zu erleiden, wird als *Hazard* $h(t)$ bezeichnet und kann als momentane Sterberate aufgefasst werden. Dieses Risiko kann sich im Zeitverlauf ändern, ist also abhängig vom Zeitpunkt t . Betrachtet man zum Beispiel die Zeit von der Medikamenteneinnahme bis zum Auftreten einer bestimmten Nebenwirkung wie z. B. Übelkeit, so ist das Risiko für Übelkeit (das Hazard) direkt nach Medikamenteneinnahme höher als das Risiko für Übelkeit einen Tag oder eine Woche später.

Hazard Ratio

Beim Vergleich zweier Gruppen kann man für beide Gruppen die Hazardfunktionen $h_1(t)$ und $h_2(t)$ bestimmen. Das Hazard Ratio ist der Quotient aus den beiden Hazard-Funktionen:

$$\text{Hazard Ratio} = \frac{h_2(t)}{h_1(t)}$$

Das Hazard Ratio ist ein Maß dafür, wie groß das Ereignisrisiko in Gruppe 2 ist, im Verhältnis zum Ereignisrisiko in Gruppe 1. Gruppe 1 ist also als Referenzgruppe anzusehen. Es gilt:

- Hazard Ratio > 1 → Ereignisrisiko in Gruppe 2 größer als in Gruppe 1
- Hazard Ratio < 1 → Ereignisrisiko in Gruppe 2 kleiner als in Gruppe 1
- Hazard Ratio ≈ 1 → Ereignisrisiko in beiden Gruppen ungefähr gleich groß.

wendet. Der Log-rank-Test untersucht, ob die Überlebenszeiten in zwei Gruppen gleich lang sind. Hierzu wird nicht nur ein bestimmter Zeitpunkt, wie zum Beispiel die 6-Monats-Überlebensrate, betrachtet, sondern der gesamte Beobachtungszeitraum. Vereinfacht kann man sagen, dass die Kaplan-Meier-Kurven miteinander verglichen werden.

In einer erweiterten Form kann der Log-rank-Test auch für den Vergleich von drei oder mehr Gruppen genutzt werden, zum Beispiel für die Gegenüberstellung der Überlebenszeiten von Patienten mit Metastasenstatus M0 versus M1 versus M2/3. In diesem Fall wird untersucht, ob in mindestens einer der Gruppen eine längere oder kürzere Überlebenszeit besteht als in den anderen Gruppen.

In der HIT'91-Studie beträgt der p-Wert des Log-rank-Tests zum Vergleich der Therapiegruppen 0,020. Zum Signifikanzniveau $\alpha = 5\%$ ist der Unterschied zwischen den Überlebenszeiten signifikant. Welche Gruppe die längeren Überlebenszeiten hat, kann man an der weiter oben liegenden Überlebenszeitkurve erkennen. Im vorliegenden Beispiel ist das die Gruppe mit Erhaltungstherapie. Patienten mit Erhaltungstherapie leben länger als Patienten mit Sandwich-Chemotherapie.

Hazard und Hazard Ratio

Kurzgefasst bezeichnet das Hazard die momentane Sterberate für eine Gruppe von Patienten. Das Hazard Ratio ist ein Quotient aus den Hazards von zwei Gruppen und gibt an, um wie viel die Sterberate in der einen Gruppe höher ist im Vergleich zu der Sterberate der anderen Gruppe. Das Hazard Ratio ist ein deskriptives Maß zum Vergleich von Überlebenszeiten zwischen zwei verschiedenen Gruppen von Patienten. Es ist wie ein relatives Risiko zu interpretieren (zum relativen Risiko: siehe *Ressing et al.* [9]) und wird in Kasten 3 näher erläutert. Beträgt das Hazard Ratio 2,3 für Patienten mit Metastasen im Vergleich zu Patienten ohne Metastasen, so ist das Sterberisiko der Patienten mit Metastasen 2,3-mal so hoch wie das der Patienten ohne Metastasen (oder auch um 130 % erhöht).

Cox-Regression

Im Folgenden soll der gleichzeitige Einfluss von mehreren Variablen auf die Überlebenszeit untersucht werden. Im Fokus stehen unter anderem:

- Therapie
- Geschlecht
- Resektionsgrad
- Metastasenstatus.

Außerdem soll der Einfluss der stetigen Variablen Alter bei Operation auf die Überlebenszeit betrachtet werden. In beiden Fällen eignet sich die Cox-Regression [3]. Die Cox-Regression bietet außerdem die Möglichkeit, einen Schätzer für die Größe des Einflusses zu erhalten. Dieser Schätzer ist durch das Hazard Ratio gegeben.

Voraussetzungen

Die Cox-Regression setzt voraus, dass das Hazard Ratio über die Zeit konstant ist (deshalb auch „proportional hazards regression“ genannt). Das ist der Fall, sobald sich das Ereignisrisiko (Hazard) der Gruppe 2 proportional zu dem von Gruppe 1 verhält (Annahme von proportionalen Hazards). Zu jedem Zeitpunkt kann zwar das Ereignisrisiko (das Hazard) unterschiedlich sein, die Unterschiede über die Zeit sollen aber in beiden Gruppen über die Zeit gleich sein. Diese Annahme ist nicht immer gerechtfertigt, lässt sich aber anhand der Kaplan-Meier-Kurven in etwa beurteilen. Wenn eine gleichmäßige Überlegenheit einer der zwei Gruppen zu erkennen ist, kann von der Annahme der proportionalen Hazards ausgegangen werden. Anschaulich gesehen ist dies der Fall, wenn die Kaplan-Meier-Kurven sich nicht kreuzen. Falls die Kurven sich kreuzen sollten, ist dies nicht ge-

	n	HR	95%-KI	p
Metastasenstatus bei Diagnose				
M0*	114			0,001
M1	33	2,11	1,13–3,94	
M2/3	40	3,06	1,76–5,33	
unbekannt	93	1,54	0,94–2,52	
Therapie				
Erhaltungstherapie*	127			0,006
Sandwich-Chemotherapie	153	1,76	1,17–2,67	
Alter bei Diagnose (in Jahren)	280	0,93	0,88–0,98	0,005

Tabelle 2 Ergebnisse der Cox-Regression für das Gesamtüberleben bei 280 Kindern mit Medulloblastom.

n: Fallzahl; HR: Hazard Ratio; 95%-KI: 95%-Konfidenzintervall für das Hazard Ratio; p: p-Wert des „likelihood ratio“-Tests; * Referenzgruppe bei kategorialen Variablen (Aus: von Hoff K et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT’91. EJC 2009; 45: 1209–17 [12]; Abdruck mit freundlicher Genehmigung des Verlags Elsevier, Oxford).

Table 2 Results of Cox regression for overall survival in 280 children with medulloblastoma.

n: no. of cases; HR: hazard ratio; 95 % CI: 95 % confidence interval for hazard ratio; p: p value of likelihood ratio test; * Control group for categorical variables. (From: von Hoff K et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT’91. EJC 2009; 45: 1209–17 [12]; printed with the kind consent of Elsevier Publishers, Oxford).

ben. *Parmar* und *Machin* [7] beschreiben, wie man die Annahme der proportionalen Hazards überprüft. Auch der Log-rank-Test setzt proportionale Hazards voraus.

Ein Beispiel für eine Situation, in der diese Annahme verletzt ist, ist folgendes: Die Überlebenszeiten von Patienten mit einer Operation sollen verglichen werden mit der von Patienten mit einer Radiotherapie anstelle der Operation. Das Risiko direkt nach der Operation zu sterben ist groß, danach sinkt das Risiko. Bei Patienten mit Radiotherapie ist das Risiko, direkt beim Beginn der Therapie zu sterben, niedrig, könnte sich aber im Laufe der Zeit erhöhen, falls die Radiotherapie nicht effektiv genug ist. Die zwei Sterberaten verhalten sich also nicht proportional zueinander.

Betrachtet man die Kaplan-Meier-Kurven der Patienten mit Metastasenstatus M1 aus der HIT’91-Studie (Grafik 2), ist zu erkennen, dass die Erhaltungstherapie

gleichmäßig überlegen ist. Somit spricht nichts gegen die Annahme der proportionalen Hazards.

Ähnlich wie bei der linearen Regression sind auch bei der Cox-Regression verschiedene Prozeduren zur Variablen-selektion möglich (siehe hierzu *Schneider* et al. [10]).

Cox-Regression am Beispiel

In der HIT’91-Studie zeigten drei Variablen einen Einfluss auf das Gesamtüberleben (Tabelle 2):

- die Therapie (binär)
- der Metastasenstatus bei Diagnose (kategorial)
- das Alter bei Diagnose (stetig).

Die Referenzgruppe für die Variable Therapie ist durch die Patienten mit Erhaltungstherapie gegeben. Ein Hazard Ratio von 1,76 lässt sich folgendermaßen interpretieren: Kinder mit Sandwich-Chemotherapie haben ein

1,76-fach so hohes Risiko zu sterben im Vergleich zu Kindern mit Erhaltungstherapie.

Der Metastasenstatus hat vier Ausprägungen:

- M0
- M1
- M2/3
- „unbekannt“ (Patienten mit unbekanntem Status sind solche, bei denen nicht klar zwischen M0 und M1 unterschieden werden konnte).

Die Referenzgruppe, mit der verglichen wird, sind die Patienten mit Metastasenstatus M0. Das Sterberisiko innerhalb der drei Gruppen M1, M2/3 und „unbekannt“ wird jeweils mit dem der Referenzgruppe M0 verglichen. Demnach werden drei Hazard Ratios ausgegeben. Kinder mit M1 haben im Vergleich zu Kindern mit M0 ein 2,11-fach so hohes Risiko zu sterben (Hazard Ratio = 2,11), das Risiko ist demnach um 111 % erhöht. Das Sterberisiko für Kinder mit

M2/3 ist 3,06-mal so hoch wie das eines Kindes mit M0. Patienten mit unbekanntem Metastasenstatus haben ein 1,54-fach so hohes Risiko zu sterben im Vergleich zu Kindern mit M0. Neben dem Hazard Ratio ist auch das Konfidenzintervall zu beachten [4]. Der Referenzwert ist hier die „1“ (= kein Effekt).

Bei einer stetigen Variable gibt das Hazard Ratio die Veränderung des Sterberisikos an, falls sich die interessierende Variable um eine Einheit erhöht, der Patient also zum Beispiel bei Diagnose ein Jahr älter ist. Mit jedem Jahr, um das ein Patient bei Diagnose älter ist, ist sein Sterberisiko um 7 % gesenkt (Hazard Ratio 0,93). Zu berücksichtigen ist, dass die gewählte Einheit der Einflussvariablen bei der Interpretation beibehalten wird (hier: Diagnosealter in Jahren, siehe Schneider et al. [10]).

Weitere wichtige Aspekte

Zeitabhängige Einflussgrößen

Bisher wurden nur Variablen betrachtet, die beim Startzeitpunkt der Überlebenszeit bekannt waren. So wurde in der HIT*91-Studie geprüft, ob Metastasen, die bei Operation des Hirntumors vorliegen, einen Einfluss auf das Überleben haben. Möchte man eine Variable untersuchen, die zum Startzeitpunkt noch unbekannt ist oder sich mit der Zeit verändert, sollte eine zeitabhängige Cox-Regression berechnet werden. Wenn man zum Beispiel wissen möchte, ob die kumulative Insulindosis bei Diabetespatienten die Zeitdauer bis zu einem kardiovaskulären Ereignis beeinflusst, darf man die kumulative Dosis nicht zum Startzeitpunkt als bekannt voraussetzen. Patienten, die länger überleben, haben im Allgemeinen eine höhere Gesamtdosis erhalten. Diese hohe kumulative Dosis hat aber nicht zum längeren Überleben geführt. Um diesen Effekt adäquat zu berücksichtigen, muss die kumulative Dosis zeitabhängig in die Cox-Regression eingehen. Eine zeitabhängige Cox-Regression durchzuführen ist sehr komplex und im Lehrbuch von Collett [2] ausführlich beschrieben.

Patienten unter Risiko

Als Patienten unter Risiko („patients at risk“) bezeichnet man die Patienten, die zu einem bestimmten Zeitpunkt noch leben. Oft wird die sich im Zeitverlauf ändernde Anzahl der Patienten unter Risiko in die Kaplan-Meier-Kurve integriert (unterhalb der Zeitachse). Da am rechten Rand der Kaplan-Meier-Kurve weniger Patienten unter Risiko sind (einige sind bereits gestorben oder zensiert), kann man mit dieser Information bestimmen, wie aussagekräftig die Kaplan-Meier-Schätzung am rechten Rand noch ist. Je weniger Patienten unter Risiko sind, desto größer wird das Konfidenzintervall des Kaplan-Meier-Schätzers.

Anzahl an Ereignissen

Um verlässliche Ergebnisse zu erzielen, muss eine angemessene Zahl an Ereignissen vorhanden sein (wichtig: hier ist nicht die Patientenzahl gemeint). Pro Variable, die in einer multivariablen Cox-Regression untersucht wird, sollten mindestens zehn Ereignisse vorliegen [8]. Bei wenigen Ereignissen können demnach nur wenige Einflussgrößen simultan untersucht werden. In der HIT*91-Studie wurden 101 Todesfälle beobachtet. Somit dürfen maximal zehn Variablen in die Cox-Regression aufgenommen werden.

Fallzahlplanung

Eine Fallzahlplanung ist sowohl für den Log-rank-Test als auch für die Cox-Regression möglich. Zusätzlich zum Signifikanzniveau und der Power, die erreicht werden soll, braucht man eine geschätzte Überlebensrate für jede Gruppe, die man vergleichen möchte beziehungsweise das geschätzte Hazard Ratio bei einer stetigen Einflussgröße [11]. Bei der Fallzahlplanung werden auch die Rekrutierungs- und Follow-up-Zeit berücksichtigt.

Zensierungen

Falls die Verteilungen der zensierten Patienten in zwei zu vergleichenden The-

rapiegruppen unterschiedlich sind, kann ein Bias in der Auswertung entstehen. Aus diesem Grunde sollte die Vollständigkeit des Follow-ups innerhalb der Therapiegruppen berichtet werden, (siehe hierzu Clark et al. [1]).

Resümee

Da Überlebenszeitdaten Zensierungen beinhalten, müssen diese immer mit dem Kaplan-Meier-Verfahren und dem Log-rank-Test ausgewertet werden. Eine Analyse anhand der Häufigkeiten der Ereignisse führt oft zu fehlerhaften Ergebnissen. Jeder Arzt sollte Kaplan-Meier-Kurven, den Log-rank-Test und Ergebnisse der Cox-Regression verstehen, weil er in der Lage sein muss, diese dem Patienten zu erklären (zum Beispiel bei der Wahl der Therapie: Hirntumorbehandlung mit Sandwich- oder Erhaltungstherapie).

Multivariable Analysen können mit der Cox-Regression durchgeführt werden. Anhand der Hazard Ratios mit Konfidenzintervallen kann man die Ergebnisse interpretieren. Leider werden die Voraussetzungen für die Cox-Regression nicht immer beachtet (zum Beispiel „proportional hazards“, zeitabhängige Variablen), deshalb sind viele publizierte Analysen oft fehlerbehaftet. Es ist wichtig, dass der Leser einer wissenschaftlichen Publikation diese Fallstricke kennt und beurteilen kann, ob die gewählte Auswertestrategie korrekt ist. 

Interessenkonflikt: Die Autoren erklären, dass kein Interessenkonflikt im Sinne der Richtlinien des International Committee of Medical Journal Editors besteht.

Korrespondenzadresse

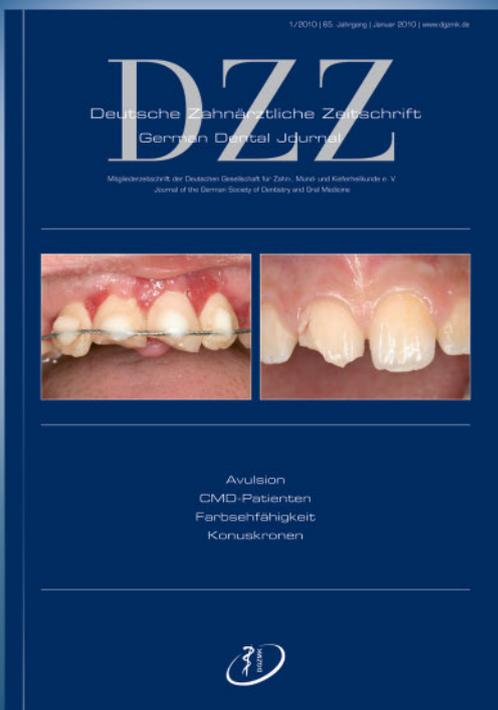
Prof. Dr. Maria Blettner
Universitätsmedizin der Johannes
Gutenberg-Universität Mainz
Institut für Medizinische Biometrie,
Epidemiologie und Informatik
Obere Zahlbacher Straße 69, 55131 Mainz
Tel.: 0 61 31 / 17 – 32 52, Fax: – 29 68
E-Mail:
maria.blettner@unimedizin-mainz.de
www.imbei.uni-mainz.de

Literatur

1. Clark TG, Altman DG, De Stavola BL: Quantification of the completeness of follow-up. *Lancet* 359, 1309–1310 (2002)
2. Collett D: *Modelling survival data in medical research*. 2nd edition. Chapman and Hall, London 2003
3. Cox DR: Regression models and life tables (with discussion). *Journal of the Royal Statistical Society (Series B)* 74, 187–200 (1972)
4. du Prel JB, Hommel G, Röhrig B, Blettner M: Confidence interval or p-value? – part 4 of a series on evaluation of scientific publications *Dtsch Arztebl Int* 106, 335–339 (2009)
5. Kalbfleisch JD, Prentice R: *The statistical analysis of failure time data*. 2nd edition. Wiley, New York 2002
6. Kaplan EL, Meier P: Nonparametric estimation from incomplete observations. *JASA* 53, 457–481 (1985)
7. Parmar MK, Machin D: *Survival analysis: a practical approach*. John Wiley and Sons, Cambridge 1995
8. Peduzzi P, Concato J, Feinstein AR, Holford TR: Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and Precision of regression estimates. *J Clin Epidemiol* 48, 1503–1510 (1995)
9. Ressing M, Blettner M, Klug SJ: Data analysis of epidemiological studies part 11 of a series on evaluation of scientific publications. *Dtsch Arztebl Int* 107, 187–192 (2010)
10. Schneider A, Hommel G, Blettner M: Linear regression analysis part 14 of a series on evaluation of scientific publications *Dtsch Arztebl Int* 107, 776–782 (2010)
11. Schumacher M, Schulgen G: *Methodik klinischer Studien*. 3rd edition. Springer, Berlin, Heidelberg, New York 2008
12. von Hoff K et al.: Long-term outcome and clinical prognostic factors in children with medulloblastoma treated in the prospective randomised multicentre trial HIT'91. *EJC* 45, 1209–1217 (2009)
13. Weiß C: *Basiswissen Medizinische Statistik*. 5th revised edition. Springer Medizin Verlag, Heidelberg 2010

WERDEN SIE LESER DER DZZ!

Seit mehr als 60 Jahren informiert die DZZ über alle Bereiche der modernen Zahnheilkunde. Überzeugen Sie sich vom hohen Praxistransfer der am meisten zitierten deutschen zahnärztlichen Zeitschrift.



Ja, ich möchte die DZZ kennenlernen. Bitte senden Sie mir kostenlos und unverbindlich ein Probeheft.

Praxisstempel/Anschrift an
Telefax +49 2234 7011-515

Vorname, Name

Straße, Postfach

Land, PLZ, Ort

E-Mail