

Überlebenszeitanalyse: Der Log-Rang-Test

– Artikel Nr. 16 der Statistik-Serie in der DMW –

Survival analysis: Log rank test

Autoren

A. Ziegler¹ S. Lange² R. Bender²

Institut

¹ Institut für Medizinische Biometrie und Statistik, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Universität zu Lübeck

² Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, Köln

Hintergrund

Der Log-Rang-Test (engl.: log rank test) gehört zu den nichtparametrischen Verfahren [1] und ist das Standardverfahren in der Überlebenszeitanalyse für Gruppenvergleiche, wie z. B. der Vergleich zweier Therapien in einer klinischen Studie. Der etwas ungewöhnliche Name Log-Rang erklärt sich daher, dass sich das Verfahren aus einem Test herleiten lässt, der auf dem Logarithmus der Ränge der Daten basiert. Der Log-Rang-Test lässt sich auch verwenden, wenn mehr als zwei Gruppen miteinander verglichen werden sollen. Hingegen werden für die gleichzeitige Untersuchung verschiedener Einflussvariablen multifaktorielle Methoden benötigt, wie z. B. das Regressionsmodell von Cox [5].

Im Folgenden wird aus Gründen der Einfachheit der Log-Rang-Test im Zwei-Gruppenvergleich beschrieben. Zur Illustration wird der Log-Rang-Test unter Verwendung der schon zuvor [6] verwendeten Daten aus Sickle-Santanello et al. [4] berechnet. Zum Abschluss der Arbeit werden Varianten und Eigenschaften des Log-Rang-Tests sowie mit ihm verwandte Verfahren diskutiert.

Der Log-Rang-Test als χ^2 -Test

Die Idee des Tests ist wie folgt: Falls die Nullhypothese richtig ist, treten die Todesfälle in zufälliger Reihenfolge unabhängig von der Gruppenzugehörigkeit auf. Ist eine der beiden Gruppen in ihrer Überlebenszeit der anderen überlegen, so werden die Todesfälle bei den Patienten dieser Gruppe später auftreten. Es ergibt sich dabei eine Abweichung zwischen den Todesfällen, die tatsächlich *beobachtet* werden, und denen, die bei zufälliger Reihenfolge der Todesfälle *erwartet* werden. Und genau die Untersuchung des Unterschieds zwischen *beobachtet* und *erwartet* ist die Basisidee der meisten so genannten χ^2 -Tests.

Um „beobachtet“ mit „erwartet“ vergleichen zu können, wird zunächst die Summe der beobachteten Todesfälle zu den einzelnen Zeitpunkten O_1 und O_2 (O für engl. „observed“) in den beiden Gruppen 1 und 2 berechnet, und anschließend in gleicher Weise die erwartete Anzahl der Ereignisse in den Gruppen 1 bzw. 2, das sind E_1 bzw. E_2 . Die Berechnung wird im Detail im Beispiel (s. u.) vorgeführt. Zum Abschluss vergleicht man den Unterschied zwischen der beobachteten und der erwarteten Anzahl der Todesfälle gemäß der Log-Rang-Teststatistik.

$$LR = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2}$$

Diese Größe ist bei hinreichend großer Zahl der Ereignisse, d. h. Todesfälle, (Faustregel mindestens 30 über beide Gruppen) annähernd χ^2 -verteilt mit 1 Freiheitsgrad (vgl. [3]).

Beispiel

Die praktische Berechnung des Log-Rang-Tests wird im Folgenden unter Verwendung der Daten von Sickle-Santanello et al. [4] zum Zungenkrebs illustriert. Gegenstand der Untersuchung sind die Daten von 80 Männern mit Zungenkrebs, die entweder ein anaploides DNA-Tumorprofil ($n = 52$) oder ein diploides DNA-Tumorprofil ($n = 28$) besitzen. Gemessen wurde die Zeit von der Diagnosestellung bis zum Tod oder die Dauer der Studienteilnahme (jeweils in Wochen). Die Überlebenszeiten aller 80 Männer mit den Eigenschaften anaploid (A) bzw. diploid (D) sind in den Spalten 2 und 3 der **Tab. 1** und grafisch als Kaplan-Meier Kurven [6] in **Abb. 1** dargestellt. Dabei bedeuten die +-Zeichen nach der Wochenzahl, dass die Beobachtung der entsprechenden Person zensiert [6] ist.

Um die erwartete Anzahl an Todesfällen in einer der beiden Gruppe zu einem Zeitpunkt $t - t$ ergibt sich aus der in der Studie verwendeten Zeit-

Schlüsselwörter

- Überlebenszeitanalyse
- Zensierte Daten
- Überlebenszeitkurven
- Log-Rang-Test

Key words

- Survival analysis
- Censored data
- Survival curves
- Log rank test

Bibliografie

DOI 10.1055/s-2007-959040
Dtsch Med Wochenschr 2007;
132: e39–41 · © Georg Thieme
Verlag KG Stuttgart · New York ·
ISSN 0012-0472

Korrespondenz

Professor Dr. rer. nat.

Andreas Ziegler

Institut für Medizinische
Biometrie und Statistik,
Universitätsklinikum Schleswig-
Holstein, Campus Lübeck,
Universität zu Lübeck
Ratzeburger Allee 160, Haus 4
23538 Lübeck
eMail
ziegler@imbs.uni-luebeck.de

Tab. 1 Berechnung der Log-Rang-Statistik für die 80 Männer mit Zungenkrebs (52 anaploides DNA-Tumorprofil, 28 diploides DNA-Tumorprofil) – Daten aus Sickle-Santonello et al. [4].

Person Nr.	Überlebenszeit t_i	Gruppe	r_i	d_i	r_{Ai}	E_{Ai}	Person Nr.	Überlebenszeit t_i	Gruppe	r_i	d_i	r_{Ai}	E_{Ai}
	0		80										
1	1	A	80	2	52	1.30	41	70	A	40	1	31	0.78
2	1	D					42	72	A	39	1	30	0.77
3	3	A	78	3	51	1.96	43	73	A	38	1	29	0.76
4	3	A					44	74+	A				
5	3	D					45	76+	D				
6	4	A	75	2	49	1.31	46	77	A	35	1	27	0.77
7	4	D					47	79+	D	34	0	26	0
8	5	D	73	2	48	1.32	48	80+	A	33	0	25	0
9	5	D					49	81+	A	32	0	24	0
10	8	D	71	1	48	0.68	50	87+	A	31	0	23	0
11	8+	D					51	87+	A	30	0	22	0
12	10	A	69	1	48	0.70	52	88+	A	29	0	21	0
13	12	D	68	1	47	0.69	53	89+	A	28	0	20	0
14	13	A	67	3	47	2.10	54	91	A	27	1	19	0.70
15	13	A					55	93	A	26	1	18	0.69
16	13	D					56	93+	A				
17	16	A	64	2	45	1.41	57	96	A	24	1	16	0.67
18	16	A					58	97+	A				
19	18	D	62	1	43	0.69	59	100	A	22	1	14	0.64
20	23	D	61	1	43	0.70	60	101+	A				
21	24	A	60	1	43	0.72	61	104	A	20	3	12	1.80
22	26	A	59	2	42	1.42	62	104	D	19			
23	26	D					63	104	D				
24	27	A	57	2	41	1.44	64	104+	A				
25	27	D					65	104+	D				
26	28	A	55	1	40	0.73	66	108+	A	15	0	10	0
27	30	A	54	3	39	2.17	67	109+	A	14	0	9	0
28	30	A					68	112	D	13	1	8	0.62
29	30	D					69	120+	A	12	0	8	0
30	32	A	51	1	37	0.73	70	129	D	11	1	7	0.64
31	41	A	50	1	36	0.72	71	131+	A	10	0	7	0
32	42	D	49	1	35	0.71	72	150+	A	9	0	6	0
33	51	A	48	1	35	0.73	73	157	A	8	1	5	0.63
34	56	D	47	1	34	0.72	74	167	A	7	1	4	0.57
35	61+	A	46	0	34	0	75	176+	D	6	0	3	0
36	62	D	45	1	33	0.73	76	181	D	5	1	3	0.60
37	65	A	44	1	33	0.75	77	231+	A	4	0	3	0
38	67	D	43	1	32	0.74	78	231+	D	3	0	2	0
39	67+	A			32		79	240+	A	2	0	2	0
40	69	D	41	1	31	0.76	80	400+	A	1	0	1	0

r_i Anzahl der Personen unter Risiko

d_i Anzahl der Ereignisse am Zeitpunkt t_i

r_{Ai} Anzahl der Personen unter Risiko in Gruppe A unmittelbar vor dem Zeitpunkt t_i

E_{Ai} Erwartete Anzahl der Ereignisse in Gruppe A am Zeitpunkt t_i

einheit, also im Beispiel zu einer bestimmten Woche – berechnen zu können, wird zunächst die Information benötigt, wie viele der Männer überhaupt noch unter Risiko r_i sind, das heißt bis unmittelbar vor t weder verstorben noch zensiert sind. Zum Monat 0 sind noch alle 80 Männer in der Stichprobe, also unter Risiko (Spalte 4). Bis unmittelbar vor der Messung zur Woche 1 sind entsprechend ebenfalls alle 80 Männer in der Stichprobe. Am Ende der Woche 1 sterben allerdings 2 Männer – einer mit anaploidem, einer mit diploidem DNA-Tumorprofil. Danach sind bis unmittelbar vor der Messung zur Woche 3 insgesamt 78

Männer unter Risiko. An diesem Untersuchungszeitpunkt sind 3 Männer ($2 \times$ Gruppe A, $1 \times$ Gruppe D) verstorben. Von Bedeutung ist es, die Zensierungen adäquat zu handhaben. In der Untersuchung zur Woche 8 ist ein Mann der Gruppe D verstorben, ein anderer zensiert. Hier wird wie bei der Kaplan-Meier Methode für die statistische Analyse angenommen, dass der zensierte Patient so lange beobachtet wurde, bis der nächste Todesfall beobachtet wird und die Zensierung erst unmittelbar danach eintritt [6]. Entsprechend sind 71 Männer unter Risiko zur Woche 8, aber nur 69 zur Woche 10.

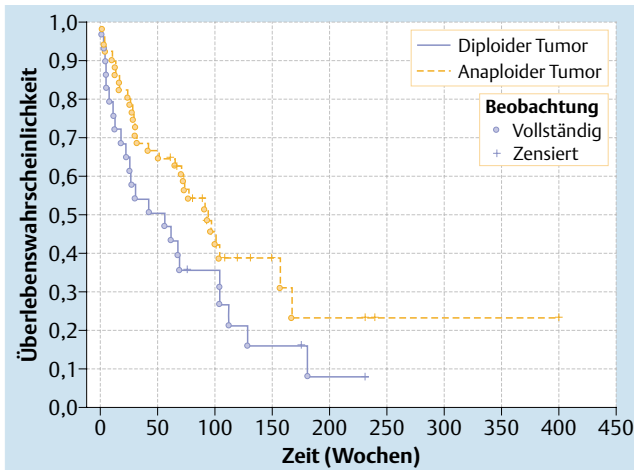


Abb. 1 Kaplan-Meier Kurven für die Überlebenszeit der 80 Zungenkrebspatienten. Es wird in orange/blau die Wahrscheinlichkeit gezeigt, dass ein Patient mit aploidem/diploidem Tumor eine Zeit (in Wochen) überlebt.

In Spalte 5 wird dann mit der Bezeichnung d_i zusammengefasst, wie viele Männer insgesamt zum Zeitpunkt t_i (unabhängig von ihrer Gruppenzugehörigkeit) tatsächlich verstorben sind. Dieses sind z. B. 2 Männer zur Woche 1 bzw. 3 Männer zur Woche 3.

Spalte 6 ähnelt sehr der Spalte 4; nur wird hier gezählt, wie viele Männer ausschließlich aus der Gruppe A zum Zeitpunkt t_i unter Risiko sind. Daher trägt die Abkürzung r_{Ai} auch den Zusatz A.

Der Anteil der Männer aus Gruppe A an allen Männern, die zu einem Zeitpunkt t_i unter Risiko stehen, ist r_{Ai}/r_i . Wenn zu einem Zeitpunkt t_i genau d_i Männer versterben, und die Gruppenzugehörigkeit für das Risiko zu versterben irrelevant ist, erwartet man, dass der Anteil der verstorbenen Männer, die der Gruppe A entstammen, gerade r_{Ai}/r_i ist. Entsprechend ist die erwartete Anzahl der Todesfälle in Gruppe A gerade $E_{Ai} = d_i \cdot r_{Ai}/r_i$.

Summiert man nun die erwartete Anzahl der Todesfälle E_{Ai} aller Patienten der Gruppe A über die einzelnen Zeitpunkte, ergibt sich ein Wert von $E_A = 36,55$. Beobachtet wurden $O_A = 31$ Todesfälle. Da in Gruppe D 22 Ereignisse beobachtet wurden, also $O_D = 22$, und die Summe aus E_A und E_D gleich der Summe aller Todesfälle sein muss, ergibt sich $E_D = O_A + O_D - E_A = 31 + 22 - 36,55 = 16,45$.

Damit erhält man als Log-Rang-Teststatistik:

$$LR = \frac{(O_A - E_A)^2}{E_A} + \frac{(O_D - E_D)^2}{E_D} = \frac{(31 - 36,55)^2}{36,55} + \frac{(22 - 16,45)^2}{16,45} = 2,715$$

Da die Log-Rang-Teststatistik annähernd χ^2 -verteilt ist mit 1 Freiheitsgrad, ergibt sich als p -Wert 0,099. Bei einem Signifikanzniveau von 5% kann somit ein unterschiedliches Risiko zwischen den beiden Gruppen nicht statistisch signifikant nachgewiesen werden.

Diskussion

Die Berechnung der Log-Rang-Teststatistik wie oben beschrieben ist recht einfach. Auch die Erweiterung auf mehr als zwei Gruppen erfolgt nach dem gleichen Schema. In diesem Fall muss man die Log-Rang-Teststatistik nur um die entsprechenden Summenterme $(O-E)^2/E$ ergänzen. Die entsprechende Teststatistik ist dann bei k Gruppen annähernd χ^2 -verteilt mit $(k-1)$ Freiheitsgraden.

Auch wenn sich der Log-Rang-Test in der Praxis einer großen Beliebtheit erfreut, gibt es eine Reihe von alternativen Verfahren, die erwähnenswert sind. Verwendet man einen etwas anderen Nenner für die Berechnung der Teststatistik, so ergibt sich die so genannte Mantel-Haenszel Version des Log-Rang-Tests (für Details siehe z. B. [2]).

Auch der generalisierte Wilcoxon-Test (engl.: generalised Wilcoxon test), auch Breslow-Test genannt, und der Tarone-Ware-Test vergleichen wie der Log-Rang-Test die Anzahl erwarteter und beobachteter Ereignisse zu jedem der Zeitpunkte miteinander. Doch tun sie dieses in unterschiedlicher Weise: Während der Log-Rang-Test alle Ereignisse gleich gewichtet, messen der generalisierte Wilcoxon-Test und der Tarone-Ware-Test die frühen Ereignisse stärker. Dieses geschieht, indem zu einem Zeitpunkt t_i anstelle der ungewichteten Größe $(O - E)$ eine gewichtete Größe $w \cdot (O - E)$ betrachtet wird. Für Details wird der interessierte Leser zum Selbststudium auf das Buch von Kleinbaum [2] verwiesen.

Die größte Power, das heißt die größte Wahrscheinlichkeit, einen tatsächlich vorhandenen Unterschied zu entdecken, hat der Log-Rang-Test, wenn die Ereignisraten in den verschiedenen Gruppen wie im Cox-Modell [5] proportional zueinander sind – ansonsten hat der generalisierte Wilcoxon-Test die größere Power. Die Frage, ob die Ereignisraten tatsächlich proportional zueinander sind, entspricht der Frage nach der Normalverteilungsannahme beim t -Test [3]. Einen schnellen Eindruck, ob die Proportionalitätsannahme erfüllt ist, erhält man durch die Kaplan-Meier Kurven [6]. Wenn diese sich kreuzen, ist eindeutige eine Abweichung von der Proportionalität gegeben. Man sollte in diesem Fall von der Verwendung des Log-Rang-Tests Abstand nehmen. Abb. 1 zeigt hier eine gute Übereinstimmung mit der Proportionalität. In onkologischen Studien könnte dieses beispielsweise dadurch entstehen, dass das Prüfpräparat eine hohe Toxizität zu Beginn aufweist, dafür aber in der Langzeit zu besseren Ergebnissen führt.

kurzgefasst

Der Log-Rang-Test ist das Standardverfahren in der Überlebenszeitanalyse für einfache Gruppenvergleiche in klinisch-therapeutischen Studien. Mit diesem nichtparametrischen Test lässt sich statistisch überprüfen, ob das Mortalitätsrisiko in zwei oder mehr Gruppen verschieden ist.

Dieser Beitrag ist eine überarbeitete Fassung aus dem Supplement Statistik aus dem Jahr 2004.

Literatur

- 1 Bender R, Lange S, Ziegler A. Wichtige Signifikanztests. Dtsch Med Wochenschr 2007; 132: e24–e25
- 2 Kleinbaum DG. Survival analysis: A self-learning text. New York: Springer, 1996
- 3 Lange S, Bender R. Was ist ein Signifikanztest? Allgemeine Aspekte. Dtsch Med Wochenschr 2007; 132: e19–e21
- 4 Sickel-Santanello BJ et al. Technical and statistical improvements for flow cytometric DNA analysis of paraffin-embedded tissue. Cytometry 1988; 9: 594–599
- 5 Ziegler A, Lange S, Bender R. Überlebenszeitanalyse: Die Cox-Regression. Dtsch Med Wochenschr 2007; 132: e39–e41
- 6 Ziegler A, Lange S, Bender R. Überlebenszeitanalyse: Eigenschaften und Kaplan-Meier Methode. Dtsch Med Wochenschr 2007; 132: e36–e38