Grafische Darstellung und Vergleich von Überlebenszeitverläufen

Ulrike Held
Horten-Zentrum, Universitätsspital, Zürich

Tabelle 1. Überlebenszeiten in Monaten der 44 Patienten.

Prednisolon Überlebenszeit (Monate)	verstorben	Plazebo Überlebenszeit (Monate)	verstorben
2	ja	2	ja
6	ja	3	ja
12	ja	4	ja
54	ja	7	ja
56	nein*	10	ja
68	ja	22	ja
89	ja	28	ja
96	ja	29	ja
125	nein	37	ja
128	nein	40	ja
131	nein	41	ja
140	nein	54	ja
141	nein	61	ja
143	ja	63	ja
145	nein	71	ja
146	ja	127	nein
148	nein	140	nein
162	nein	146	nein
168	ja	158	nein
173	nein	167	nein
181	nein	182	nein

* Lost to follow-up.

Ulrike Held

Die Autorin erklärt, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Beitrag hat.

In klinischen Studien und in der medizinischen Forschung möchte man häufig die Zeit bis zum Auftreten eines bestimmten Ereignisses analysieren. Dieses Ereignis kann ein positives (z.B. Geburt, Heilung einer bestimmten Erkrankung), aber auch ein negatives (z.B. Tod, Auftreten eines Rezidivs) sein. Unabhängig vom tatsächlichen Ereignis bezeichnet man die statistische Methode hierfür Überlebenszeitanalyse.

Überlebensdaten sind dadurch gekennzeichnet, dass bei einer bestimmten gegebenen Studienlänge nicht unbedingt bei allen beobachteten Individuen das Ereignis schon eingetreten ist, so dass man für diese zwar keine exakte Dauer, aber einen Mindestwert angeben kann. Ausserdem kann es noch sein, dass der Überlebensstatus von einigen Individuen nicht bekannt ist, weil diese vorzeitig aus der Studie ausgeschieden sind *(Lost to follow-up)*, oder an einer anderen Erkrankung gestorben sind (konkurrierendes Ereignis). Für alle diese Individuen nennt man die Beobachtungszeit zensiert, was bedeutet, dass die Beobachtungszeit endete, bevor das Ereignis eingetreten war.

In Tabelle 1 finden sich die Überlebenszeiten von 44 Patienten mit chronisch aktiver Hepatitis [1]. Diese Patienten wurden in zwei Gruppen randomisiert: die eine Hälfte erhielt Prednisolon, die andere Hälfte Plazebo. Bei einer gewöhnlichen statistischen Auswertung kann man beim Vorliegen von wenigen fehlenden Werten diese Patienten manchmal einfach aus der Analyse ausschliessen. Dies ist allerdings bei der Analyse von Überlebenszeiten nicht angezeigt, da die Individuen mit fehlender exakter Überlebenszeit trotzdem noch einen wichtigen Beitrag zu den Daten liefern: nämlich, dass sie mindestens solange überlebt haben (zensierte Beobachtungszeiten). Würde man diese Individuen aus der Analyse ausschliessen, bekäme man ein verfälschtes Ergebnis, weil tendenziell eher lang überlebende Patienten ausgeschlossen würden. Auch darf man die zensierte Beobachtungszeit nicht als echte Beobachtungszeit betrachten. Stattdessen müssen statistische Methoden verwendet werden, die der speziellen Datenstruktur Rechnung tragen.

Kaplan-Meier-Kurve

Typischerweise werden die Verläufe von Überlebenszeiten mit einer sogenannten Kaplan-Meier-Kurve dargestellt. Diese Darstellung ermöglicht in unserem Beispiel einen optischen Vergleich zwischen der Gruppe von Patienten mit Prednisolon und mit Plazebo, wobei der speziellen Datenstruktur Rechnung getragen wird, dass es Individuen mit und ohne Ereignis gibt. Für die Darstellung als Kaplan-Meier-Kurve werden *bedingte* Überlebenswahrscheinlichkeiten für jedes Zeitintervall berechnet. Die Wahrscheinlichkeit ist bedingt, weil sich die Überlebenswahrscheinlichkeit nur auf diejenigen Patienten bezieht, die zu Beginn des untersuchten Intervalls noch am Leben waren. In unserem Beispiel haben in der Prednisolon-Gruppe 21/22 = 95,5% der Patienten die ersten zwei Monate überlebt (Zeitintervall 1). Im zweiten Zeitintervall von Monat 2–6 haben 20/21 = 95,2% überlebt. Von Monat 6–12 (Zeitintervall 3) haben 19/20 = 95,0% überlebt, und so weiter. Um nun die Überlebenswahrscheinlichkeit bis Monat 12 zu berechnen, werden diese drei bedingten Wahrscheinlichkeiten miteinander multipliziert, und man kommt auf 0,955 × 0,952 × 0,950 = 0,864, also 86,4%. Dieser Wert ist in Abbildung 1 genau bei Monat 12 eingezeichnet. Wir

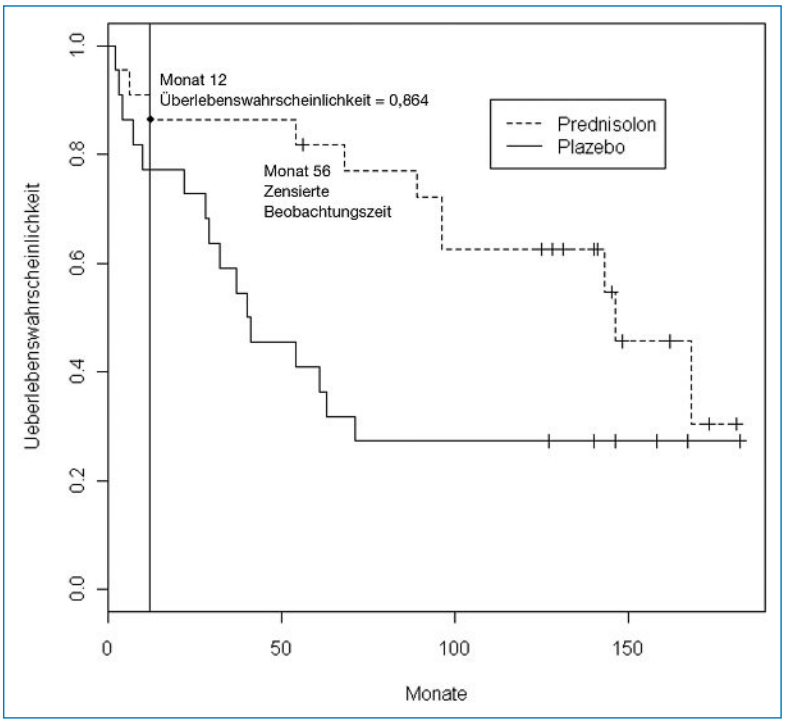

Abbildung 1
Kaplan-Meier Überlebenszeitkurven für die Patienten der Prednisolon- und der Plazebo-Gruppe.

Tabelle 2. Überlebenswahrscheinlichkeiten in der Prednisolon-Gruppe.

Intervall	Anzahl Patienten	Todesfälle	Zensiert	Überlebenswahrscheinlichkeit
2	22	1	0	0,955
6	21	1	0	0,909
12	20	1	0	0,864
54	19	1	0	0,818
56	18	0	1	0,818
68	17	1	0	0,770
89	16	1	0	0,722
96	15	2	0	0,626
125	13	0	1	0,626
128	12	0	1	0,626
131	11	0	1	0,626
140	10	0	1	0,626
141	9	0	1	0,626
143	8	1	0	0,548
145	7	0	1	0,548
146	6	1	0	0,456
148	5	0	1	0,456
162	4	0	1	0,456
168	3	1	0	0,304
173	2	0	1	0,304
181	1	0	1	0,304

haben für die Prednisolon-Gruppe alle Überlebenswahrscheinlichkeiten in Tabelle 2 eingetragen.
Stellt man diese Überlebenswahrscheinlichkeiten für beide Gruppen in einer Grafik dar, so kann man die Zahlen direkt miteinander vergleichen. Die Überlebenszeitkurven machen genau zu den Zeitpunkten «Sprünge», an denen Ereignisse beobachtet werden. Zusätzlich sind in der Grafik die zensierten Beobachtungen als vertikale Striche eingezeichnet: wir wissen von diesen Patienten nur, dass sie mindestens solange überlebt haben. Im Allgemeinen berechnet der Computer diese Treppenfunktion, und auch *Varianzen* und *Konfidenzintervalle* können hierfür angegeben werden. Vergleicht man nun die beiden Kurven miteinander, stellt man fest, dass die Patienten in der Plazebo-Gruppe eine geringere Überlebenswahrscheinlichkeit haben als die Prednisolon-Patienten, denn die Kurve der Plazebo-Patienten liegt immer unterhalb derer der Prednisolon-Patienten.
Allerdings stecken in der obigen Darstellung einige Annahmen, die möglicherweise schwierig zu prüfen sind. Erstens unterstellen wir, dass die Patienten mit den zensierten Beobachtungszeiten dieselben Verläufe haben wie die unzensierten Patienten, was in der Realität nicht leicht überprüft werden kann. Zweitens gehen wir davon aus, dass alle Patienten in einer Behandlungsgruppe die gleichen Überlebenswahrscheinlichkeiten haben, egal ob sie früh oder spät in die Studie aufgenommen wurden. Dies kann bei langjährigen Beobachtungsstudien problematisch werden, da sich eventuell die Behandlungsmöglichkeiten im Laufe der Zeit verbessert haben.
Drittens nehmen wir an, dass das Ereignis genau zu dem spezifizierten Zeitpunkt auftritt. Falls das Ereignis wie in unserem Beispiel «Tod» ist, dann ist dies eindeutig. Falls aber das Ereignis z.B. «Auftreten eines Rezidivs» in einer Krebsstudie ist, und das Ereignis bei einer Routineuntersuchung festgestellt wurde, dann ist der genaue Zeitpunkt des Ereignisses eben nicht bekannt, sondern nur das Zeitintervall. Solange die Untersuchungen sehr regelmässig stattfinden, kann man statistisch-methodisch damit umgehen, sind die Untersuchungsintervalle allerdings sehr unterschiedlich lang, führt dies zu einem *Bias*, nämlich einer Überschätzung der tatsächlichen Zeit bis zum Auftreten des Rezidivs und somit der rezidivfreien Überlebenswahrscheinlichkeit.

Log-rank-Test

Man möchte nun den Vergleich zwischen den beiden Behandlungsgruppen nicht nur optisch vollziehen, sondern mithilfe eines geeigneten statistischen Tests. Ein *t-Test* für den Zwei-Gruppen-Vergleich ist in diesem Fall nicht möglich, da wir dann ignorieren müssten, dass bei einem Teil der Patienten das Ereignis noch nicht eingetreten ist. Stattdessen verwendet man bei Überlebenszeiten typischerweise den sogenannten Log-rank-Test, der wie ein Anpassungstest funktioniert und für jedes Zeitintervall die beobachtete Anzahl Todesfälle mit der erwarteten Anzahl Todesfälle vergleicht, unter der Nullhypothese, dass kein Unterschied zwischen den Kurven vorliegt.
In unserem Beispiel ist der p-Wert des Log-rank-Tests gerade 0,031, das bedeutet also, wir können einen

statistisch signifikanten Unterschied in der Überlebenswahrscheinlichkeit zwischen den zwei Gruppen nachweisen.
Da der Log-rank-Test mit den zensierten Beobachtungszeiten genauso umgeht wie die Kaplan-Meier-Methode, gelten auch dieselben Annahmen. Also Zensierung muss unabhängig von der Prognose sein, Überlebenswahrscheinlichkeiten sind unabhängig davon, ob ein Individuum früh oder spät rekrutiert wurde, und das Ereignis muss genau zu dem spezifizierten Zeitpunkt aufgetreten sein und eben nicht nur irgendwann davor. Wenn die Annahmen nicht genau erfüllt sind, dann ist das am problematischsten, wenn sie sich in den Gruppen unterscheiden. Vor der Verwendung des Log-rank-Tests sollten die Überlebenszeitkurven immer graphisch dargestellt werden, denn es kann auch möglich sein, dass sie sich kreuzen. In diesem Fall sollten – falls genügend Daten vorhanden sind – separate Tests in den Intervallen berechnet werden, in denen sich die Kurven nicht kreuzen.

Cox-Regression

Wir haben nun gelernt, wie man Überlebenszeitkurven graphisch darstellen kann und wie man zwei oder mehr Kurven miteinander vergleichen bzw. auf einen Unterschied testen kann. Darüber hinaus möchte man vielleicht Überlebensdaten mit einer/mehreren Einflussgrössen im Rahmen eines Regressionsansatzes zusammenbringen und vor allem deren Effekt schätzen. Hierzu verwendet man typischerweise die *Cox-Regression*, oder *Cox-proportional-hazards*-Regression, die im nächsten Kapitel beschrieben werden soll.

Glossar

Bedingte Wahrscheinlichkeit

Man nennt die Wahrscheinlichkeit für das Eintreten von Ereignis A unter der Bedingung, dass das Ereignis B schon eingetreten ist, die bedingte Wahrscheinlichkeit von A gegeben B, oder P(A|B). Im Zusammenhang mit der Analyse von Überlebenszeiten interessiert man sich für die Wahrscheinlichkeit, ein bestimmtes Zeitintervall zu überleben, falls man zu Beginn des Intervalls noch gelebt hat. In anderem Zusammenhang möchte man z.B. die bedingte Wahrscheinlichkeit schätzen, an Lungenkrebs (A) zu erkranken, wenn man Nichtraucher (B) ist.

Bias

Von Bias spricht man beim Schätzen von populationsbezogenen Grössen. Dieser Begriff bezeichnet systematische Fehler, die zur Verzerrung von Schätzwerten und Tests führt. Man unterscheidet verschiedene Biasquellen, die wichtigsten sind Selektionsbias und Publikationsbias.

Konfidenzintervall

Ein Konfidenzintervall oder Vertrauensbereich für einen geschätzten Parameter einer Verteilung ist ein Intervall, in dem plausible Werte des wahren unbekannten Parameters liegen. Bei Wiederholung des Experiments liegt der wahre Parameter in 95% der Fälle innerhalb dieses Intervalls (95%-Konfidenzintervall). Sehr breite Konfidenzintervalle deuten entweder auf eine hohe Variabilität in den Daten hin oder auf eine sehr kleine Stichprobe.

Lost to follow-up

Teilnehmer in einer wissenschaftlichen Studie sind «Lost to follow-up», wenn sie für den Beobachter nicht mehr zur Verfügung stehen und der Kontakt abgebrochen ist. Dies kann verschiedene Gründe haben: die Teilnehmer können verstorben sein, sie können umgezogen sein, sie können aufgrund der Nebenwirkungen nicht mehr an der Studie teilnehmen wollen, usw. Es ist sehr wichtig, den Gründen für «Lost to follow-up» nachzugehen, denn sie können in der Schlussbeurteilung eine Verzerrung bewirken.

t-Test

Der (2-Stichproben) t-Test prüft, ob die Mittelwerte zweier Stichproben gleich sind. Im 1-Stichproben-Fall vergleicht man den Mittelwert einer Stichprobe mit einem vorgegebenen Wert. Zusätzlich zu den Gruppenmittelwerten geht in die Teststatistik auch die Quadratwurzel aus der Varianz ein.

Varianz

Die Varianz ist ein Mass für die Streuung einer Variablen.

Korrespondenz:
Dr. rer. nat. Ulrike Held
Horten-Zentrum
Universitätsspital Zürich
Postfach Nord
CH-8091 Zürich
Ulrike.Held@usz.ch

Weiterführende Literatur

– Kirk AP, Jain S, Pocock S, Thomas HC, Sherlock S. Late results of the Royal Free Hospital prospective controlled trial of prednisolone therapy in hepatitis B surface antigen negative chronic active hepatitis. Gut. 1980;21:78–83.
– Altman D, Bland M. Statistics Notes: Time to event (survival) data. BMJ. 1998;317:468–9.
– Bland M, Altman D. Statistics Notes: Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317:1572.
– Bland M, Altman D. Statistics Notes: The logrank test. BMJ. 2004;328:1073.
– Held L, Rufibach C, Seifert B. Einführung in die Biostatistik. 4. Auflage. Zürich: Abteilung Biostatistik, Institut für Sozial- und Präventivmedizin der Universität Zürich; Juli 2009. http://www.biostat.uzh.ch.
– Hüsler J, Zimmermann H. Statistische Prinzipien für medizinische Projekte. 4. Auflage. Bern: Huber-Verlag; 2006.
– Kreienbrock L, Schach S. Epidemiologische Methoden. 4. Auflage. München: Elsevier Verlag; 2005.
– R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2008. ISBN 3–900051–07–0, URL http://www.R-project.org.