

Überlebenszeitanalyse: Eigenschaften und Kaplan-Meier Methode

– Artikel Nr. 15 der Statistik-Serie in der DMW –

Survival analysis: Properties and Kaplan-Meier method

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

Schlüsselwörter

- ▶ Überlebenszeitanalyse
- ▶ Zensierte Daten
- ▶ Überlebenszeitkurve
- ▶ Kaplan-Meier Methode

Key words

- ▶ Survival analysis
- ▶ Censored data
- ▶ Survival curve
- ▶ Kaplan-Meier method

Bibliografie

DOI 10.1055/s-2007-959038
Dtsch Med Wochenschr 2007;
132: e36–e38 · © 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

Überlebenszeitanalyse



In vielen kontrollierten klinischen Studien wird die Wirksamkeit therapeutischer Maßnahmen anhand der Zeit bis zum Auftreten eines bestimmten Ereignisses beurteilt. Dabei können diese Ereignisse positiv sein, wie z. B. die Entlassung aus einer Klinik oder das Eintreten einer Schwangerschaft bei Paaren mit Kinderwunsch. Sie können neutral sein, wie z. B. das Ende der Stillzeit, oder sie können negativ sein, wie z. B. der Tod oder das Auftreten einer Krankheit [1]. Unabhängig von der Wertung des Ereignisses wird in der Medizin ganz allgemein von Überlebenszeitanalyse (engl.: survival analysis) gesprochen. In der Demografie oder den Sozialwissenschaften werden üblicherweise die Begriffe Sterbetafelanalyse, Verweildaueranalyse, Verlaufsdatenanalyse oder Ereigniszeitanalyse (engl.: life table methods, failure time data analysis, lifetime data analysis) verwendet.

In der Medizin gibt es eine Vielzahl von Anwendungsmöglichkeiten: So wird z. B. in onkologischen Therapiestudien die Überlebenszeit oder die Zeit bis zum Auftreten eines Rezidivs unter zwei verschiedenen Therapien miteinander verglichen. Mitunter soll ebenfalls eine Prognose für die erwartete Überlebenszeit eines Patienten abgegeben werden, z. B. wie lange es im Mittel dauern wird, bis eine Patientin mit Kinderwunsch schwanger wird.

In all den oben beschriebenen Situationen ist das Charakteristische der Überlebenszeitanalyse, dass die Zielvariable nicht zu einem festen Zeitpunkt erhoben werden kann, d. h. dass es zu Beginn einer Studie unbekannt ist, wann das Ereignis eintritt. Darüber hinaus muss am Ende des Beobachtungszeitraums das Ereignis nicht eingetreten sein; dann wird von einer zensierten Beobachtungszeit gesprochen, oder kurz: Einer zensierten Beobachtung. Zensierung kann auch

dadurch entstehen, dass der Patient in der Beobachtung verloren geht, also „lost to follow-up“ ist. Eine Zensierung ist auch durch das Eintreten eines konkurrierenden Risikos (engl.: competing risk) möglich, wenn z. B. der Patient durch einen fremdverursachten Verkehrsunfall ver stirbt und nicht an seinem Tumor, wobei das interessierende Ereignis der tumorbedingte Tod ist.

Sehr häufig werden im Rahmen von Überlebenszeitstudien Patienten in einem bestimmten Zeitraum, also nicht zu einem bestimmten Zeitpunkt, rekrutiert und über einen anderen bestimmten Zeitraum hinaus mindestens nachbeobachtet. Beträgt die Rekrutierungsphase z. B. zwei Jahre und die Nachbeobachtungsphase drei Jahre, so dauert die Studie insgesamt fünf Jahre. Dabei wird die zuerst rekrutierte Person fünf Jahre nachbeobachtet und die zuletzt rekrutierte Person drei Jahre, falls bei ihr innerhalb des Beobachtungszeitraums das Zielereignis nicht eingetreten sein sollte. Aufgrund der kürzeren Nachbeobachtungszeit hat die zuletzt rekrutierte Person eine geringere Wahrscheinlichkeit, dass bei ihr das Ereignis eintritt. Daher ist eine wesentliche Annahme bei vergleichenden Überlebenszeitstudien, z. B. im Rahmen von epidemiologischen Kohortenstudien, dass die Prognose für die verschiedenen Patienten in der Studie gleich ist, also unabhängig vom Rekrutierungszeitpunkt. Bei randomisierten Studien ist das Problem weniger gravierend, da sich die Nachbeobachtungszeiten zwischen den Gruppen aufgrund der Randomisierung nicht wesentlich unterscheiden werden. Dieselbe Annahme macht man üblicherweise auch für Patienten, die in der Nachbeobachtung verloren gehen.

Bei Verletzung dieser Annahmen kann es leicht zu Fehlschlüssen im Hinblick auf mögliche kausale Zusammenhänge kommen. Beispielsweise könnte eine Überschrift der Boulevard-Presse

lauten [3]: „Alarm! 95% aller Bundesliga-Fußballer werden keine 75 Jahre alt“. Diese Aussage impliziert einen Vergleich mit der (Allgemein-) Bevölkerung, von der wir wissen, dass die Mehrzahl der Menschen älter als 75 Jahre wird. Genau bei diesem (impliziten) Vergleich wird die Unabhängigkeitsannahme von Zensurierung und Prognose verletzt: Betrachtet man das Gründungsjahr 1963 der Bundesliga und nimmt man einmal an, dass der älteste Spieler damals 35 Jahre alt war, ist oder wäre der älteste Spieler von damals heute (im Jahr 2007) höchstens 78 Jahre alt. Es bestand somit für keinen Bundesliga-Fußballer die Chance, bis heute z. B. 80 Jahre alt geworden zu sein, d. h. die Beobachtungen wurden für alle spätestens bei einem Alter von 78 Jahren zensiert. Damit haben aber gerade die Zensierten eine besonders gute Prognose. Umgekehrt müssen die Nicht-Zensierten eine besonders schlechte Prognose gehabt haben, weil sie schon in jungen Lebensjahren verstorben sind. In der impliziten Vergleichsgruppe der Allgemeinbevölkerung existiert ein solcher Zensurierungsmechanismus natürlich nicht, denn hier sind sowohl früh (vor dem Alter von 75 Jahren) Verstorbene als auch noch lebende 100-Jährige vertreten.

Das hier in einem nicht-medizinischen Kontext beschriebene Phänomen wird als „lead-time bias“ bezeichnet und ist insbesondere bei der Beurteilung der Effektivität von Screeningmaßnahmen bedeutsam, wenn die Zeit von Diagnosestellung bis zum Tod betrachtet wird. Durch das Screening wird der Diagnosezeitpunkt für asymptomatisch Erkrankte nach vorne verschoben, die ohne Screening erst beim Auftreten von Symptomen diagnostiziert würden, beispielsweise im Mittel 6 Monate später. Somit haben beim Vergleich von gescreenten mit nicht gescreenten Erkrankten die gescreenten per se einen „Vorsprung“ (engl.: lead) in ihrer Überlebenszeit von eben 6 Monaten. Einen ähnlichen „Vorsprung“ hat auch die Allgemeinbevölkerung in dem fiktiven Vergleich mit den Bundesliga-Fußballern (s. o.). Allgemein gilt, dass bei Vergleichen von Überlebenszeiten, wenn unterschiedlich lange Beobachtungszeiten und Zensierungsmuster bestehen, eine Reihe von Annahmen für eine korrekte Schlussfolgerung notwendig sind: a) die Prävalenz der Risikofaktoren bleibt konstant, b) die Charakteristika der Population unter Risiko bleiben konstant und c) die Prognose des Überlebens bleibt über die Zeit konstant. Unglücklicherweise sind diese Annahmen häufig nicht haltbar und generell nicht testbar.

Kaplan-Meier Methode

Tab. 1 zeigt die Daten der Überlebenszeit von männlichen Zungenkrebspatienten mit diploidem DNA-Profil aus einer Studie zum Vergleich der Überlebenszeit bei anaploidem und diploidem DNA-Profil [4]. In dem Beobachtungszeitraum verstarben 22 der Patienten; bei sechs Patienten war die Beobachtungszeit zensiert. Ziel ist es nun, trotz der Zensurierung für jedes Zeitintervall die Wahrscheinlichkeit für das Überleben eines Patienten zu berechnen. Hierfür wird die Kaplan-Meier Methode [2] verwendet. Die entscheidende Idee dieser Methode ist, dass die Ereignisse die Beobachtungsintervalle definieren, und nicht, dass die Beobachtungsintervalle fest vorgegeben sind. Ein neues Zeitintervall wird dadurch definiert, dass ein Patient verstirbt. Für jedes Zeitintervall wird die bedingte Wahrscheinlichkeit berechnet, dass der Patient das Zeitintervall überlebt, falls er schon bis zum Beginn des neuen Zeitintervalls überlebt hat. Die

Tab. 1 Überlebenszeit (Wochen) von 28 Männern mit Zungenkrebs und diploidem DNA-Tumorprofil – Daten aus Sickle-Santanello et al. [4].

	Verstorben		Zensiert
1	18	69	8
3	23	104	67
4	26	104	76
5	27	112	104
5	30	129	176
8	42	181	231
12	56		
13	62		

Gesamtwahrscheinlichkeit dafür, einen bestimmten Zeitpunkt zu überleben, lässt sich dann darstellen als Produkt der entsprechenden bedingten Wahrscheinlichkeiten.

Dieses wird im Folgenden anhand der Daten illustriert. Von den 28 Männern war einer nach einer Woche verstorben; überlebt haben 27 von 28, also: $27/28 = 0,964$. Von den 27 Männern zu Beginn der zweiten Woche verstarb einer. Daher ist die bedingte Wahrscheinlichkeit, die zweite Woche zu überleben, wenn die erste Woche überlebt wurde, $26/27 = 0,963$. Die Gesamtwahrscheinlichkeit, die zweite Woche zu überleben, ist $0,964 \cdot 0,963 = 0,929$. So kann für alle weiteren Zeitpunkte verfahren werden: nach der vierten Woche waren noch 25 Patienten am Leben; in der fünften Woche verstarben zwei, so dass die bedingte Wahrscheinlichkeit, die fünfte Woche zu überleben, $23/25 = 0,920$ beträgt. Die Gesamtwahrscheinlichkeit, die fünfte Woche zu überleben, beträgt $0,964 \cdot 0,963 \cdot 0,962 \cdot 0,920 = 0,821$ (siehe [Abb. 1](#)). Die berechneten Überlebenszeiten lassen sich mit der Kaplan-Meier Grafik darstellen ([Abb. 1](#)).

Doch wie ist mit einer Zensurierung umzugehen, die erstmalig in der achten Woche auftritt? Hier wird für die Berechnung die wichtige Annahme gemacht, dass die Daten eines zensierten Patienten bis zum Ende des nächsten Todesfalls vorliegen und die Zensurierung erst unmittelbar danach eintritt. Daher wird der Patient, dessen Beobachtung nach der achten Woche zensiert ist, in die Berechnung der Überlebenszeit für die achte Woche ein-

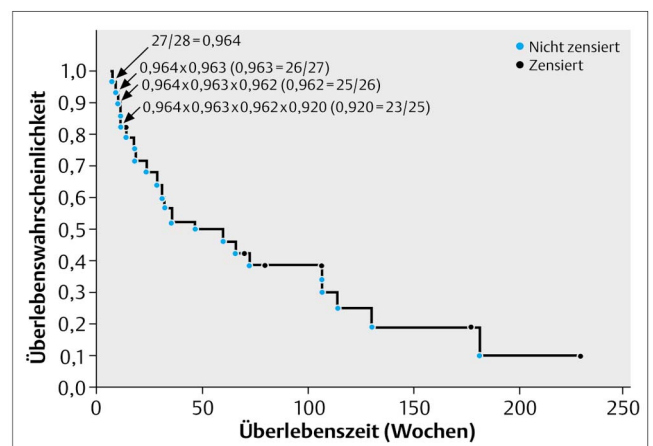


Abb. 1 Kaplan-Meier Kurve für die Überlebenszeit der 28 Zungenkrebspatienten mit diploidem Tumor. Es wird die Wahrscheinlichkeit gezeigt, dass ein Patient eine Zeit (in Wochen) überlebt.

bezogen. Aber dieser Patient wird natürlich nicht für die Berechnung der bedingten Wahrscheinlichkeit herangezogen, die neunte Woche zu überleben, wenn die achte Woche überlebt wurde.

Die Annahme, dass eine zensierte Beobachtung bis zum nächsten Todesfall vorhanden ist, kann man **Abb. 1** entnehmen: bei den vier weiteren Patienten mit zensierten Daten verläuft die Kaplan-Meier Kurve horizontal über den Zensierungszeitpunkt hinaus weiter.

Aus der Kaplan-Meier Kurve lassen sich problemlos spezielle Überlebensraten, wie z. B. die 1-Jahres-Überlebensrate ablesen. Diese beträgt in dem Beispiel 49% (Ablesen bei 52 Wochen). Umgekehrt lassen sich aus der Grafik auch spezielle Überlebenszeiten ablesen. So beträgt die mediane Überlebenszeit im Beispiel 42 Wochen (Ablesen bei Überlebenswahrscheinlichkeit 0,5).

Über die Kaplan-Meier Kurve hinaus werden weitere statistische Verfahren benötigt, mit denen es möglich ist, formale Tests in zwei Gruppen durchzuführen. Der entsprechende Log-Rang-Test (engl.: log rank test) wird in einem weiteren Artikel beschrieben [5]. Aufgrund des engen ideellen Zusammenhangs zur multiplen logistischen Regression wird das populäre Cox-Modell, mit dem Adjustierungen für Kovariablen möglich sind, ebenfalls in einer separaten Arbeit beschrieben [6].

kurzgefasst

Mit Hilfe der Überlebenszeitanalyse lässt sich die Wahrscheinlichkeit berechnen, dass ein Ereignis bis zu einem bestimmten Zeitpunkt eintritt. Dieses ist mit der Kaplan-Meier Methode auch dann möglich, wenn nicht alle Patienten identische Beobachtungszeiträume haben. Die Kaplan-Meier Kurve ist hilfreich zum Ablesen spezifischer Überlebensraten oder Überlebenszeiten.

Dieser Beitrag ist eine überarbeitete Fassung aus dem Supplement Statistik aus dem Jahr 2002.

Literatur

- 1 Altman DG, Bland JM. Time to event (survival) data. *BMJ* 1998; 317: 468–469
- 2 Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Ass* 1958; 53: 457–481
- 3 Krämer W. So lügt man mit Statistik. Frankfurt: Campus, 1991
- 4 Sickle-Santanello BJ, Farrar WB, DeCenzo JF 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. Der Log-Rang-Test. *Dtsch Med Wochenschr* 2007; 132: e39–e41
- 6 Ziegler A, Lange S, Bender R. Das Cox-Modell. *Dtsch Med Wochenschr* 2007; 132: e42–e44