

Kaplan–Meier & Co.

Ereigniszeitdaten werden in der Regel mit einer Kaplan–Meier–Kurve dargestellt. Ob sich zwei oder mehr Kaplan–Meier–Kurven signifikant unterscheiden, kann mit dem Log–Rang–Test ermittelt werden.

Die Cox–Regression erlaubt es, den Einfluss mehrerer Faktoren auf die Überlebensfunktion zu ermitteln.

Roland Müller-Waldeck

Der Kaplan–Meier–Schätzer (KM)

Der KM nutzt alle vorhandenen Daten, um die Überlebensraten in der Stichprobe darzustellen, also zu zeigen, bei wie vielen Patienten in der Stichprobe bis zu einem bestimmten Zeitpunkt kein Endpunkt eingetreten ist. Die mathematische Funktion dafür wird Überlebensfunktion genannt und als $S(t)$ bezeichnet.

Für den KM wird ein Beobachtungsintervall durch ein beobachtetes Ereignis vorgegeben, also wenn ein Patient in der Stichprobe einen Endpunkt erleidet. Die Wahrscheinlichkeit für jeden einzelnen Studienteilnehmer, diesen Zeitpunkt zu überleben, kann man leicht berechnen:

(Anzahl Patienten unter Risiko – Anzahl Patienten mit Zielereignis) / Anzahl Patienten unter Risiko

Diese Rechnung führt man für jeden Ereigniszeitpunkt durch und multipliziert die Ergebnisse miteinander, um die Wahrscheinlichkeit zu errechnen, einen Zeitpunkt t nach der Randomisierung zu überleben. Die Ergebnisse für jeden Ereigniszeitpunkt werden so verbunden, dass eine stufenförmige Grafik entsteht, die Kaplan–Meier–Kurve.

Für zensierte Daten trifft man die Annahme, dass die Daten bis zum nächsten Ereignis nach der Zensierung vorliegen, nach dem Ereignis nicht mehr. Zensierte Patienten erscheinen also immer nur als Patient unter Risiko, niemals als Ereignis, denn der Endpunkt wird bei ihnen nicht beobachtet. Endet eine Stu-

die mit einer zensierten Beobachtung, erreicht die Kurve nicht den Nullpunkt der y-Achse. Endet die Studie mit einer nicht zensierten Beobachtung, endet die Kurve auf dem Nullpunkt. Zensierte Daten sollten in der Kaplan–Meier–Kurve erkennbar sein, sonst ist die Aussagekraft sehr eingeschränkt.

Der KM ist, wie der Name schon sagt, ein statistischer Schätzwert für den tatsächlichen Wert in der Grundgesamtheit. Daher können für die verschiedenen Ereigniszeitpunkte Konfidenzintervalle errechnet werden, die ebenso wie die Punkte der Ereigniszeiten stufenförmig miteinander verbunden werden.

Besonderes Augenmerk sollte bei der Interpretation von KM–Kurven auf den Endpunkt gelegt werden:

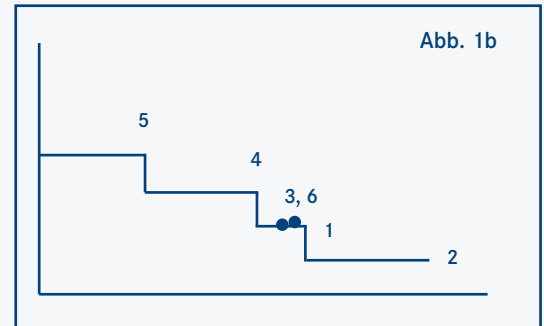
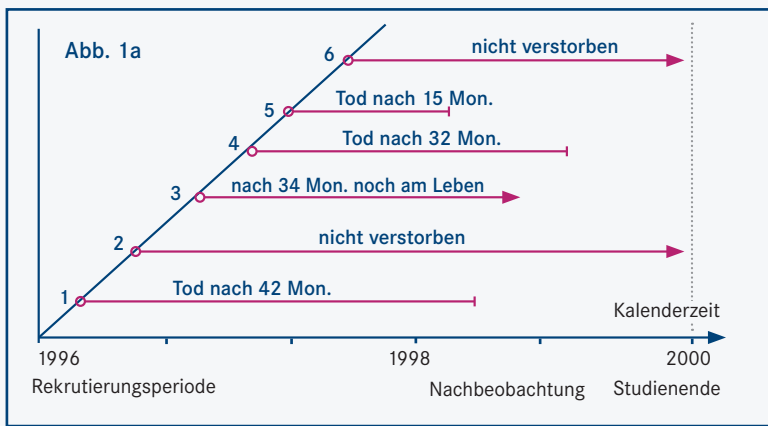
- Bei overall-survival-Kurven ist der Endpunkt der Tod durch jedwedes Ereignis.
- Bei disease-free-survival-Kurven ist der Endpunkt ein Rückfall. Diese Kurven verlaufen grundsätzlich niedriger als vergleichbare overall-survival-Kurven, weil die Patienten

früher einen Rückfall erleiden, als sie sterben.

- Bei progression-free-survival-Kurven ist Krankheitsprogression der Endpunkt. Mit diesem Endpunkt lässt sich ermitteln, wie gut eine Erkrankung auf die Behandlung anspricht.
- Bei disease-specific-survival-Kurven ist der Tod durch die untersuchte Erkrankung der Endpunkt. Diese Kurven können missverständlich sein, da sie grundsätzlich höher verlaufen als die Kurven für overall-survival und disease-free-survival.

Um eine KM–Kurve beurteilen zu können, muss man zunächst den beobachteten Endpunkt und die Einheit auf der x-Achse klären. Kurven mit vielen kleinen Stufen haben viele Teilnehmer und geben die tatsächlichen Verhältnisse besser wieder als Kurven mit wenigen Teilnehmern. Sind viele Daten zensiert, sollte man die Ursache dafür suchen. Weil die Kurve nach dem ersten zensierten Datum nur noch eine Näherung ist, ist die Kurve umso unsicherer, je mehr zensierte Daten einfließen. Kurven, in denen zensierte Daten nicht kenntlich gemacht sind, sollte man mit Vorsicht beurteilen. Weil die KM–Kurve weit rechts nur noch auf wenigen Teilnehmern basiert, sollte auch dieser Kurvenverlauf mit Vorsicht betrachtet werden.

Literatur: Martin Schumacher, Gabi Schulgen, Methodik klinischer Studien, 3. Aufl. 2008, Springer. Elisa T. Lee, Osca T. Go, Survival analysis in Public health Research. Annu. Rev. Public health. 1997; 18:105-134. Jason T. Rich et al. A Practical Guide to understanding Kaplan–Meier Curves. Otolaryngol Head Neck Surg. 2010; 143(3):331-336. Isabella Zwiener et al. Überlebenszeitanalyse. Deutsches Ärzteblatt 2011; 108(10):163-169. A. Ziegler et al. Überlebenszeitanalyse: Eigenschaften und Kaplan–Meier Methode. Dtsch Med Wochenschr 2007;132:e36-e38. Ziegler et al. Überlebenszeitanalyse: Der Log–Rang–Test. Dtsch Med Wochenschr 2007;132:e39-e41. A. Ziegler et al. Überlebenszeitanalyse: Die Cox–Regression. Dtsch Med Wochenschr 2007;132:e42-e44. Kate Bull et al. Tutorial in Biostatistics. Survival Analysis in observational studies. Statistics in Medicine, Vol 16,1041-1074 (1997)



Ablauf einer klinischen Studie mit Kaplan-Meier-Kurve

Abb. 1a: Patient 1 wird als erster rekrutiert und stirbt nach 42 Monaten, Patient 5 stirbt schon nach 15 Monaten, von Patient 3 ist nur bekannt, dass er nach 34 Monaten noch gelebt hat (zensierte Daten), Patienten 2 und 6 haben am Ende der Nachbeobachtungszeit noch gelebt (zensierte Daten). Daraus ergibt sich die schematische Kaplan-Meier-Kurve.

Abb. 1b: Die Zahlen geben die Patientennummern wieder: Patienten 5, 4 und 1 erleiden einen Endpunkt, zensiert sind die Daten von Patienten 3, 6 und 2. Vor Studienende zensierte Daten sind mit einem schwarzen Punkt auf der Linie gekennzeichnet, mit Studienende zensierte Daten lassen sich daran erkennen, dass die Kurve den Nullpunkt nicht erreicht. (Nach Schulgen, 2008)

Der Log-Rang-Test

Der Log-Rang-Test ist ein Signifikanztest, mit dem man ermitteln kann, ob sich zwei KM-Kurven signifikant unterscheiden.

Die Ereigniszeitdaten zweier KM-Kurven, die man vergleichen möchte, kann man in Form von Vierfeldertafeln darstellen (Abb. 2), deren Randsummen man als fest betrachtet.

Die Nullhypothese lautet, dass die Ereigniszeiten in beiden Gruppen gleich verteilt sind. Wenn das so ist, dann müssen in der Gruppe 1 proportional so viele Teilnehmer verstorben sein wie in Gruppe 1 und 2 zusammen (beide Gruppen unterscheiden sich nicht). Die Anzahl erwarteter Ereignisse für Gruppe 1 wäre also die Anzahl der Ereignisse zum Zeitpunkt t dividiert durch die Gesamtzahl an Patienten unter Risiko (durchschnittliches Risiko für beide Gruppen).

Dieser Wert wird nun multipliziert mit der Anzahl Patienten unter Risiko in Gruppe 1 und ergibt die erwartete

Anzahl von Ereignissen in Gruppe 1 = $(d_i/n_i) * n_{1i}$. Aus den Differenzen zwischen erwarteten und beobachteten Häufigkeiten lässt sich ein p-Wert ermitteln. Ist dieser Wert unter 5%, gilt der Unterschied zwischen Gruppe 1 und Gruppe 2 als signifikant (die Nullhypothese wird abgelehnt). Auch mehr als 2 Gruppen lassen sich mit diesem Test miteinander vergleichen.

Die Cox-Regression

Regressionsanalyse bei der Überlebenszeitanalyse

Oft stellt sich die Frage, wie mehrere Faktoren die Entwicklung einer Krankheit beeinflussen, oder es soll der Einfluss von Confoundern abhängig von der Zeit ermittelt werden. Das erfordert eine multivariate Analyse. Mit Hilfe der Cox-Regression lässt sich der Einfluss mehrerer Variablen auf die Überlebenszeit ermitteln. Weil das relative Risiko als Maß dafür, wie ein Faktor die Mortalität beeinflusst, die Zeitachse nicht

berücksichtigt, hat sich der Hazard (Risiko, dass zum Zeitpunkt t ein bestimmtes Ereignis eintritt) etabliert, der das Ereignisrisiko abhängig von der Zeit angibt:

$$h(t) = \frac{\text{Anzahl Personen mit Endpunkt im Zeitintervall } t_1 \text{ bis } t_2}{\text{Anzahl Personen unter Risiko zum Zeitpunkt } t_1 * \text{Breite des Zeitintervalls}}$$

Mit Hilfe der Cox-Regression können für mehrere Risikofaktoren gleichzeitig die Hazard Ratios (HR) bezogen auf eine Vergleichsgruppe berechnet werden. Ist der HR gleich 1, haben beide Gruppen das gleiche Risiko, ist der Wert größer als 1, ist das Risiko für den untersuchten Faktor höher als in der Vergleichsgruppe (HR=1,30 bedeutet ein 30% erhöhtes Risiko), ist der Wert kleiner als 1, gilt das Umgekehrte.

Die Cox-Regression setzt voraus, dass die Risiken der untersuchten Faktoren proportional zueinander sind, was bedeutet, dass sich die KM-Kurven nicht kreuzen dürfen. Ist das der Fall, muss ein anderes Verfahren eingesetzt werden. ■

	Anzahl Ereignisse an t_i	Anzahl ereignisfreier Patienten an t_i	Anzahl Patienten unter Risiko kurz vor t_i
Gruppe 1	d_{1i}	$n_{1i} - d_{1i}$	n_{1i}
Gruppe 2	d_{2i}	$n_{2i} - d_{2i}$	n_{2i}
	d_i	$n_i - d_i$	n_i

Log-Rang-Test

In Abb. 2 sind die Daten von Zeitpunkt t_1 in einer Vierfeldertafel zusammengefasst. Aus den Vierfeldertafeln lässt sich für alle Ereigniszeiten ein p-Wert ermitteln.

(Nach Schulgen, 2008)