1. Analiza przeżycia – podstawowe informacje

Czas do wystąpienia zdarzenia

Cenzurowanie

Ucinanie

Szacowanie p-stwa przeżycia

Analiza przeżycia

- Dla szczególnego rodzaju danych *ciągłych*:

  czasu do wystąpienia zdarzenia

  - (nieujemny) czas od pewnej *chwili początkowej* do *zdarzenia*

- <u>Przykłady</u>:
  - czas przeżycia (zdarzenie=zgon)
  - czas przeżycia bez objawów (zdarzenie=nawrót choroby/zgon)
  - czas przeżycia bez progresji (zdarzenie=progresja choroby/zgon)

- Inne określenia: czas do wystąpienia niepowodzenia, czas niepowodzenia, czas przeżycia

- Szczególna cecha danych: ***cenzurowanie obserwacji***

Cenzurowanie

- Nie obserwujemy dokładnego czasu zdarzenia T^*

- Trzy głowne rodzaje cenzurowania:
  - ***prawostronne***: obserwujemy dolną granicę czasu ($T^* > C$)
    - Przykład: brak wznowy guza przed końcem próby klinicznej
  - ***lewostronne***: obserwujemy górną granicę czasu ($T^* < C$)
    - Przykład: wznowa guza przed pierwszym badaniem kontrolnym
  - ***przedziałowe***: obserwujemy przedział czasu ($C_L < T^* < C_U$)
    - Przykład: wznowa guza między dwoma badaniami kontrolnymi

Inne przykłady danych cenzurowanych

- Cenzurowanie może dotyczyć wszystkich danych ciągłych
- Często wynika z ograniczeń aparatury pomiarowej

- Testy przemysłowe
  - np. siły potrzebnej do zniszczenia jakiegoś przedmiotu
  - może być większa niż maksymalna w przyrządzie testującym

- Mikromacierze, spektrometria masowa, itp.:
  - saturacja intensywności sygnału
  - lub wartość sygnału poniżej granicy mierzalności

Obserwacja w próbach klinicznych (1)

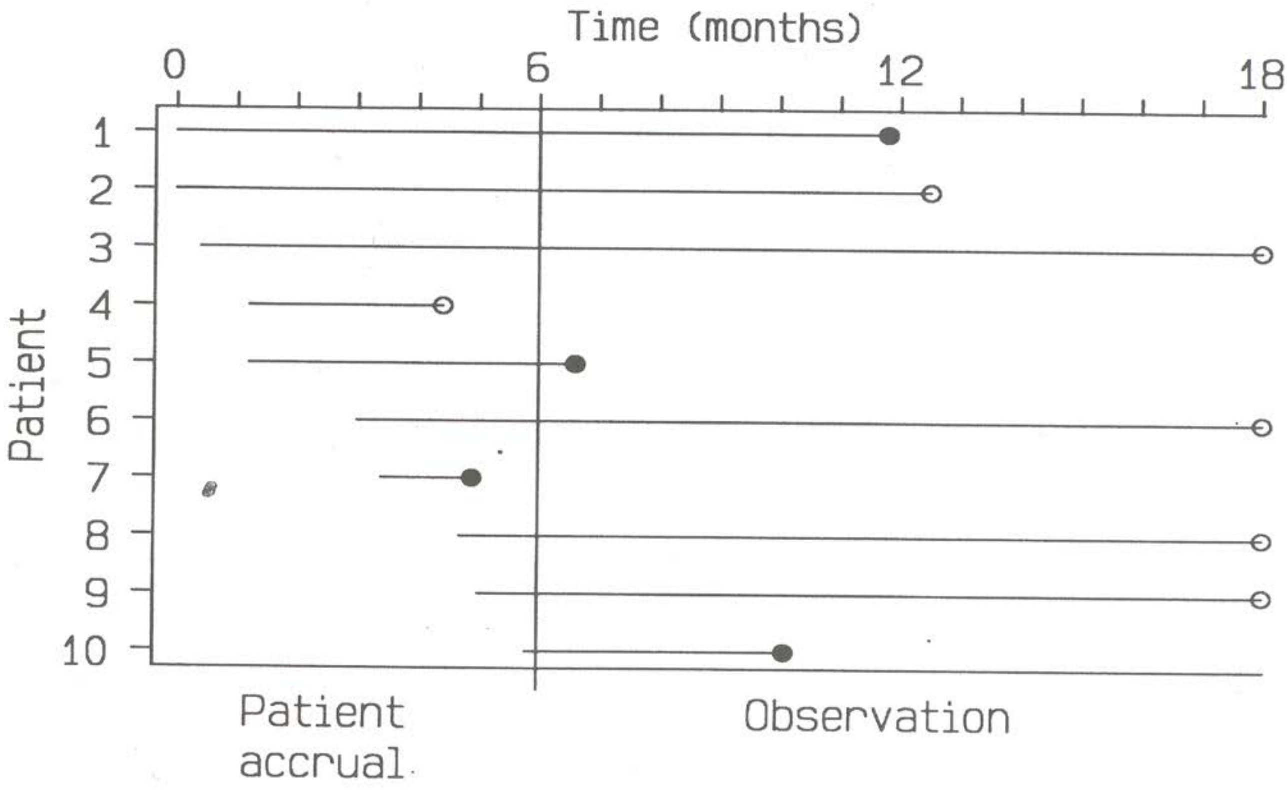


Figure 13.1 Diagram showing patients entering a study at different times and the observation of known (●) and censored (○) survival times.

Obserwacja w próbach klinicznych (2)

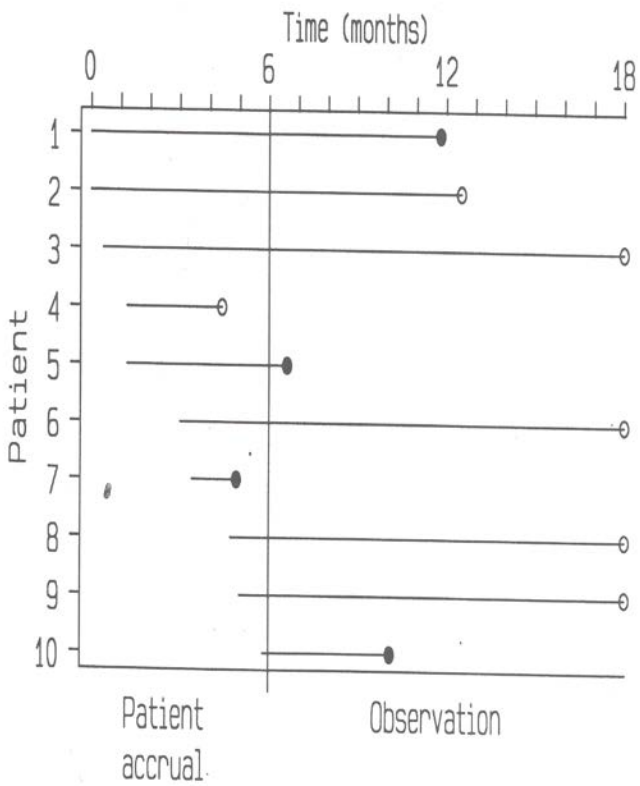


Figure 13.1 Diagram showing patients entering a study at different times and the observation of known (●) and censored (○) survival times.

Pacjent	Randomizacja (czas w mies.)	Ostatnia obserwacja (mies.)	Zmarł?
1	0.0	11.8	tak
2	0.0	12.5	nie
3	0.4	18.0	nie
4	1.2	4.4	nie
5	1.2	6.6	tak
6	3.0	18.0	nie
7	3.4	4.9	tak
8	4.7	18.0	nie
9	5.0	18.0	nie
10	5.8	10.1	tak

Prawostronnie cenzurowany czas do wystąpienia zdarzenia

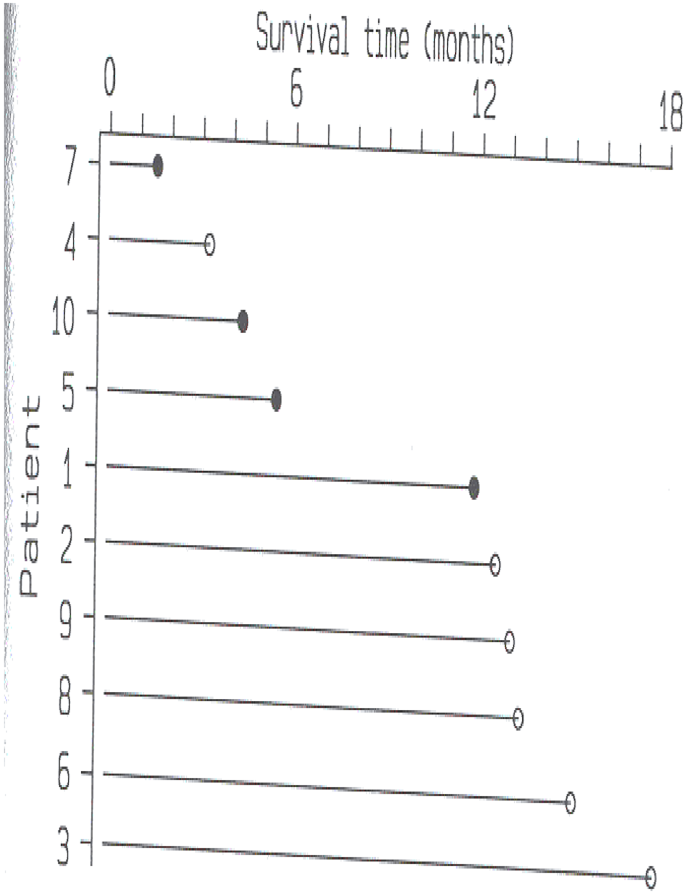


Figure 13.2 Figure 13.1 reorganized to correspond to method of analysis.

Pacjent k	Czas przeżycia (mies.) t_k	Wsk. zgonu (1=tak) d_k
1	11.8	1
2	12.5	0
3	17.6	0
4	3.2	0
5	5.4	1
6	15.0	0
7	1.5	1
8	13.3	0
9	13.0	0
10	4.3	1

Problem związany z cenzurowaniem

- Chcemy oszacować p-stwo zgonu przed upływem 6 miesięcy
  - 6 pacjentów przeżyło, 3 zmarło przed upływem 6 miesięcy
  - O 1 pacjencie (# 4) wiemy jedynie, że przeżył(a) 3.2 miesiąca (obserwacja cenzurowana prawostronnie).
- Jak oszacować to p-stwo?
  - *4/10* → przeszacowanie (# 4 mógł przeżyć 6 miesięcy)
  - *3/10* → niedoszacowanie (# 4 mógł umrzeć przed upływem 6 miesięcy)
  - *3/9* → ? + strata informacji (wiemy, że # 4 przeżył(a) 3.2 miesiąca)

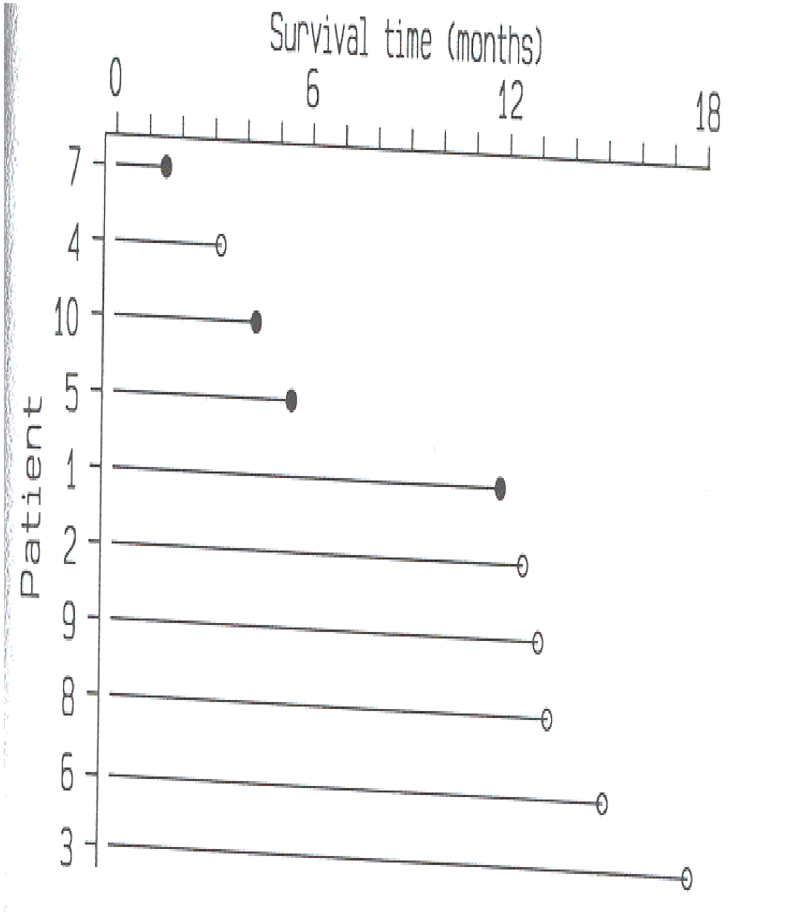


Figure 13.2 Figure 13.1 reorganized to correspond to method of analysis.

Formalizm i notacja (1)

- *Czas do wystąpienia zdarzenia*: zm. losowa $T^* \geq 0$
  - Będziemy zakładać, że T^* jest ciągła
- *Cenzurowany prawostronnie czas do wystąpienia zdarzenia*: $T=\min(T^*, C)$, gdzie zm. losowa $C \geq 0$
- *Wskaźnik cenzurowania*: $\delta = \mathrm{I}(T^* \leq C)$
  - Obserwujemy pary (T, δ)

Formalizm i notacja (2)

- Rozkład p-stwa T^* z funkcją gęstości $f(t)$
- *Funkcja przeżycia* $S(t) = P(T^* \geq t)$
  - Dla T^* ciągłej, $S(t)=1-F(t)$, gdzie $F(t)$ - dystrybuanta
  - $S(0)=1$
- *Funkcja hazardu* $$\lambda(t) = \lim_{h \to 0^+} \frac{P\left(t \leq T^* < t+h \mid T^* \geq t\right)}{h}$$
- *Funkcja skumulowanego hazardu* $$\Lambda(t) = \int_0^t \lambda(u)\,du$$

Funkcje hazardu: przykłady

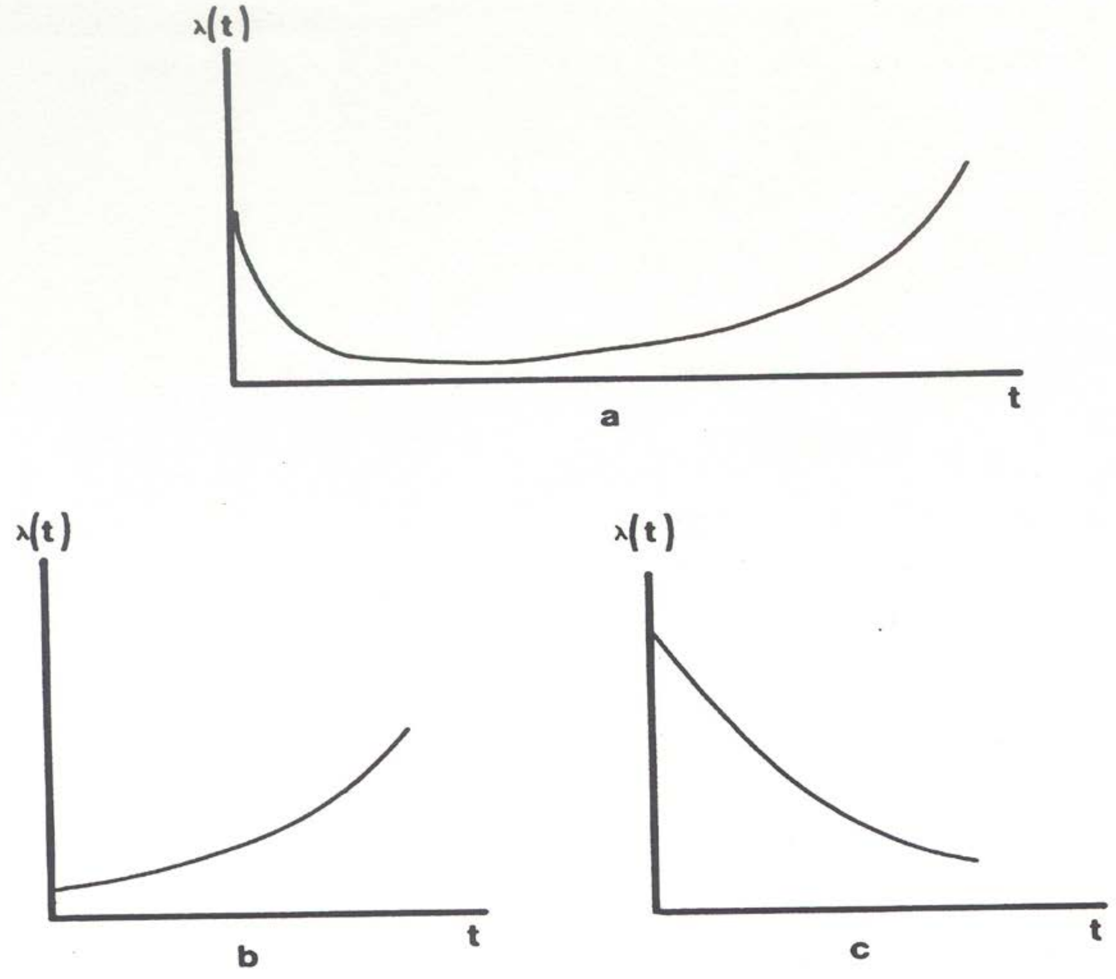


Figure 1.1 Some types of hazard functions: (a) hazard for human mortality; (b) positive aging; (c) negative aging.

Zależności (1)

$$
\begin{aligned}
\lambda(t) &= \lim_{h \to 0^+} \frac{P\left(t \le T^* < t+h \mid T^* \ge t\right)}{h} \\
&= \lim_{h \to 0^+} \frac{P\left(t \le T^* < t+h\right)}{h} \frac{1}{P\left(T^* \ge t\right)} \\
&= \lim_{h \to 0^+} \frac{F\left(t+h\right) - F\left(t\right)}{h} \frac{1}{S\left(t\right)} \\
&= \frac{f(t)}{S\left(t\right)}
\end{aligned}
$$

Zależności (2)

$$\lambda(t) = \frac{f(t)}{S(t)} = -\frac{d}{dt}\{\log S(t)\}$$

$$\Downarrow$$

$$S(t) = e^{-\int_0^t \lambda(u)du} \equiv e^{-\Lambda(t)}$$

$$\Downarrow$$

$$\Lambda(t) = -\log S(t)$$

Przykładowe rozkłady: wykładniczy

$$\lambda(t) \equiv \lambda$$

$$\Lambda(t) = \lambda t$$

$$S(t) = e^{-\lambda t}$$

$$f(t) = \lambda e^{-\lambda t}$$

- $E(T^*)=1/\lambda$; $Var(T^*)=1/\lambda^2$; $med(T^*)=\ln 2/\lambda$
- Brak pamięci: $P(T^* \geq t+u \mid T^* \geq t)=P(T^* \geq u)$
- $E(T^*-t \mid T^* \geq t)=1/\lambda$
  - Stała *oczekiwana pozostała długość życia*

Przykładowe rozkłady: Weibull

$$\lambda(t) \equiv \lambda p(\lambda t)^{p-1}$$

$$\Lambda(t) = (\lambda t)^{p}$$

$$S(t) = e^{-(\lambda t)^{p}}$$

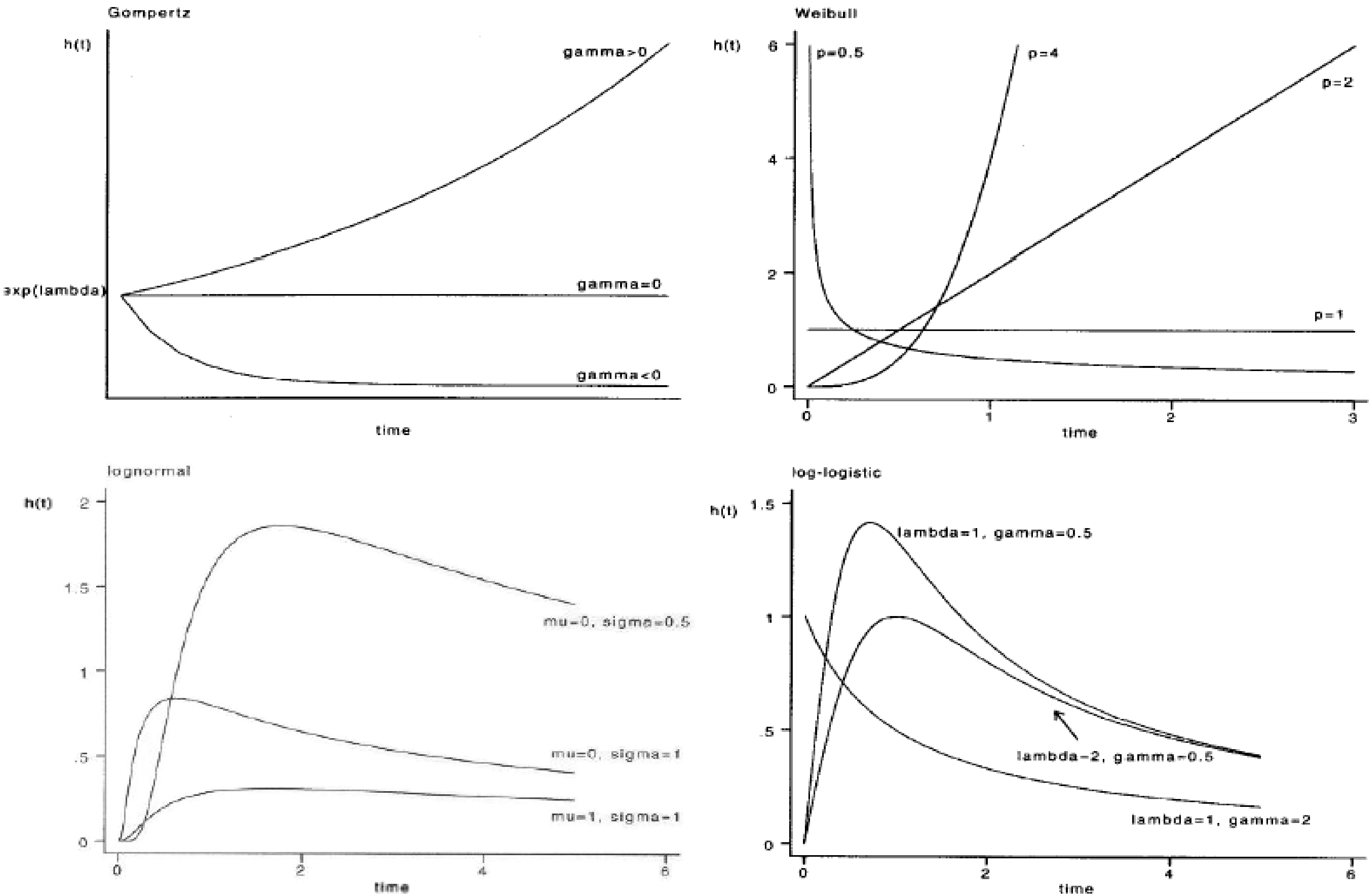
$$f(t) = \lambda p(\lambda t)^{p-1} e^{-(\lambda t)^{p}}$$

- *λ – parametr skali*; *p – parametr kształtu*
- Dla *p*=1, rozkład wykładniczy

Inne używane rozkłady

- T^* ~ *Gompertz*
- T^* ~ *(Uogólniony) Gamma*
- $\ln T^*$ ~ *Normalny* (T^* ~ *log-normalny*)
- $\ln T^*$ ~ *Logistyczny* (T^* ~ *log-logistyczny*)

Funkcje hazardu



Cenzurowanie prawostronne typu I

- n par $(T_1^*, C_1), (T_2^*, C_2), \ldots, (T_n^*, C_n)$
- Obserwujemy $T_i = \min(T_i^*, C_i)$, gdzie C_i = const
- Testowanie niezawodności
  - Jednoczesne rozpoczęcie obserwacji n jednostek
  - Obserwujemy czasy zdarzeń, jeśli są krótsze niż C
  - Dla dłuższych czasów zdarzeń, obserwujemy C

Progresywne cenzurowanie prawostronne typu I (próba kliniczna)

- Obserwacja rozpoczynana w różnych chwilach (kalendarzowych) $t_{0,i}$
- T_i^* iid, niezależne od chwil rozpoczęcia obserwacji
- Zakończenie obserwacji w <u>ustalonym</u> czasie C (kalendarzowym)
- $T=\min(T_i^*, C - t_{0,i})$

Cenzurowanie prawostronne typu II

- n par $(T_1^*, C_1), (T_2^*, C_2), \ldots, (T_n^*, C_n)$

- Obserwujemy $T_i = min(T_i^*, C_i)$, gdzie $C_i = T_{(r)}^*$
  - $T_{(r)}^*$ - *r-ta statystyka porządkowa*

- Testowanie niezawodności
  - Jednoczesne rozpoczęcie obserwacji n jednostek
  - Obserwacja do uzyskania r-tego zdarzenia

Proste losowe cenzurowanie prawostronne

- n par $(T_1^*, C_1), (T_2^*, C_2), \ldots, (T_n^*, C_n)$
- C_i są iid, niezależne od $T_1^*, \ldots, T_n^*$
- Obserwujemy $T_i = min(T_i^*, C_i)$

Cenzurowanie niezależne

- Cenzurowanie jest niezależne, jeśli

$$\lim_{h\to 0^+}\frac{P\left(t \le T^* < t+h \mid T^* \ge t\right)}{h} = \lim_{h\to 0^+}\frac{P\left(t \le T^* < t+h \mid T^* \ge t, Y(t)=1\right)}{h}$$

  Y(*t*)=1 jeśli do chwili *t* nie wystąpiło zdarzenie ani cenzurowanie (jednostka pozostaje *narażona na ryzyko* zdarzenia).

- Interpretacja: jednostka cenzurowana w chwili *c* jest reprezentatywna dla wszystkich innych narażonych na ryzyko zdarzenia w chwili *c*.

Cenzurowanie niezależne (cd.)

- Definicję można rozszerzyć tak, by uwzględniała zmienne objaśniające *X*=*x*:

$$\lim_{h\to 0^+}\frac{P\left(t\le T^* < t+h \mid T^*\ge t, x\right)}{h}=\lim_{h\to 0^+}\frac{P\left(t\le T^* < t+h \mid T^*\ge t, Y(t)=1, x\right)}{h}$$

- Interpretacja: jednostka cenzurowana w chwili *c* jest reprezentatywna dla wszystkich innych narażonych na ryzyko zdarzenia w chwili *c* i *o tych samych wartościach zmiennych objaśniających.*

Cenzurowanie nie-informatywne

- Załóżmy T_i^* iid z rozkładu z funkcją gęstości $f(t;\theta)$
- Niech C_i iid z rozkładu z funkcją gęstości $g(t;\theta,\varphi)$.
- Cenzurowanie jest nie-informatywne, jeśli

$$g(t;\theta,\varphi) \equiv g(t;\varphi)$$

- Interpretacja: cenzurowanie nie daje informacji o parametrach rozkładu czasów zdarzeń.

Niezależne i nie-informatywne cenzurowanie

- W dalszych rozważaniach będziemy zakładać, że cenzurowanie jest niezależne i nie-informatywne.

Ucinanie

- Jednostka włączona do próbki jeśli $a<T^*<b$
  - Jeśli $a=-\infty$, *ucinanie prawostronne*
  - Jeśli $b=+\infty$, *ucinanie lewostronne*

- W istocie oznacza próbkowanie z rozkładu warunkowego.

Ucinanie: przykład

- Lui et al. (1986)
- T^* = czas od zakażenia HIV do AIDS
- Daty diagnozy AIDS i transfuzji z HIV z rejestru
- Pierwsza diagnoza: 01/06/1982; ostatnia 31/12/1984

- *Ucinanie prawostronne* dla transfuzji po 01/06/1982
  - z badania wykluczone osoby z T^* dłuższym niż 2.5 roku
- *Prawo- i lewostronne* dla transfuzji przed 01/06/1982
  - z badania wykluczone osoby z T^* krótszym niż do 01/06/1982 i dłuższym niż do 31/12/1984

Ucinanie i cenzurowanie

- Ucinanie jest czasem mylone z cenzurowaniem.

- Ucinanie oznacza, że próbkujemy z niepełnej populacji, tzn. z rozkładu warunkowego.
  - Dla ucinania lewostronnego, szacujemy $S(t)/S(a)$.

- Cenzurowanie oznacza, że dla niektórych jednostek w próbie mamy tylko częściową informację dotyczącą czasu T^* (np. $T^* > C$).
  - Niezależnie od tego, czy mamy ucinanie, czy nie

Formalizm i notacja (3)

- Funkcja wiarogodności dla cenzurowania prawostronnego:

$$L(\theta,\varphi) = L(\theta)L^*(\theta,\varphi)$$

gdzie

$$L(\theta) = \prod_{j=1}^{n} f\left(t_j;\theta\right)^{\delta_j} S\left(t_j;\theta\right)^{1-\delta_j} = \prod_{j=1}^{n} \lambda\left(t_j;\theta\right)^{\delta_j} S\left(t_j;\theta\right)$$

a $L^*(\theta,\varphi)$ zależy od cenzurowania (parametr φ).

  - Przy niezależnym cenzurowaniu, $L(\theta)$ jest *częściową* funkcją wiarogodności.
  - Jeśli cenzurowanie jest dodatkowo nie-informatywne, $L(\theta)$ jest pełną funkcją wiarogodności, bowiem wówczas $L^*(\theta,\varphi) \equiv L^*(\varphi)$.

Główne mechanizmy cenzurowania w próbach klinicznych

- Trzy główne mechanizmy:
  - ***Cenzurowanie administracyjne***: związane z końcem próby
  - ***Wycofanie z próby***: pacjent nie chce kontynuować leczenia w próbie
  - ***Strata z obserwacji***: pacjent przestaje zgłaszać się na wizyty kontrolne

Cenzurowanie nieinformatywne

- Cenzurowanie administracyjne zazwyczaj spełnia warunki niezależności i nieinformatywności

- Wycofania/straty z obserwacji są <u>potencjalnie problemem</u>:
  - mogą wynikać np. z progresji choroby lub zgonu

- Należy ograniczać ich liczbę
  - w ostateczności, próbować ustalać przyczynę

Dane cenzurowane prawostronnie

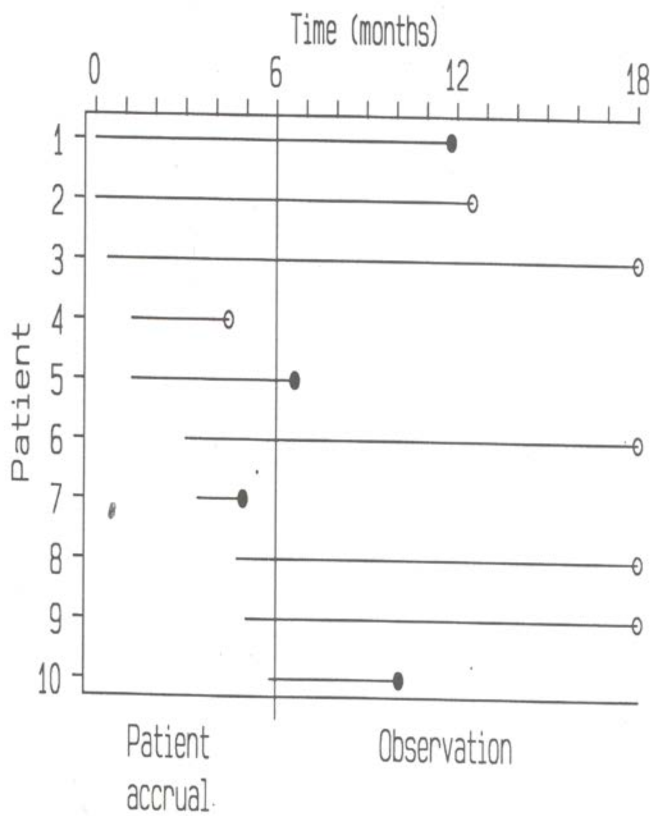


Figure 13.1 Diagram showing patients entering a study at different times and the observation of known (•) and censored (○) survival times.

Pacjent	Randomizacja (mies.)	Ostatnia obserwacja (mies.)	Zmarł ?	Czas przeżycia (mies.)	Wskaźnik zgonu (1=tak)
1	0.0	11.8	yes	11.8	1
2	0.0	12.5	no	12.5	0
3	0.4	18.0	no	17.6	0
4	1.2	4.4	no	3.2	0
5	1.2	6.6	yes	5.4	1
6	3.0	18.0	no	15.0	0
7	3.4	4.9	yes	1.5	1
8	4.7	18.0	no	13.3	0
9	5.0	18.0	no	13.0	0
10	5.8	10.1	yes	4.3	1

Analiza przeżycia: metodologia

- Występowanie obserwacji cenzurowanych wymaga użycia specjalnych metod
  - posługiwanie się np. średnią próbkową lub odchyleniem próbkowym jest <u>błędem</u>

- Tradycyjnie podstawowym parameterem jest prawdopodobieństwo, że zdarzenie nie wystąpi przed upływem czasu *t:*

S(t) = P(czas do wystąpienia zdarzenia jest dłuższy niż t)

  - *funkcja przeżycia* (zależność *p-stwa przeżycia* od czasu)
  - $S(t) = P(T \geq t)$

Estymator Kaplana-Meiera (1)

- Idea:
  - by przeżyć *t* jednostek czasu, najpierw trzeba przeżyć pierwszych *t-1* jednostek, a potem jeszcze jednostkę *t*

- Symbolicznie:

$$S(t)=S(t-1) \cdot P(\text{przeżycie } t\text{. jednostki czasu})$$

  - zakładamy, że $S(0)=1$

Estymator Kaplana-Meiera (2)

- Obserwowane czasy $t_1, t_2, \ldots, t_n$
- Uporządkowane czasy zdarzeń: $t_{(1)}, t_{(2)}, \ldots, t_{(d)}$
  - $d \leq n$
- n_j = *liczebność zbioru ryzyka dla* $t_{(j)}$
  - liczba jednostek obserwowanych tuż przed $t_{(j)}$
- d_j = *liczba zdarzeń dla* $t_{(j)}$

$$\hat{S}(t) = \prod_{t_{(j)} \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Przykład: choroba lokomocyjna (1)

Burns, *Aviat Space Environ Med* (1984)

- 21 osób poddanych 2-godz „kołysaniu”…
- … o częstości 0.167 Hz i przyśpieszeniu 0.111 G
- Czas do pierwszych torsji ☹
- Dwie osoby zażądały przerwania doświadczenia (tylko! ☺ ☺)

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla *t* R_t	Zdarzenia dla *t* D_t	Przeżycie (0 zdarzeń) *t* $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-*t*) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas *t*	Zbiór ryzyka dla *t* R_t	Zdarzenia dla *t* D_t	Przeżycie (0 zdarzeń) *t* $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-*t*) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla *t* R_t	Zdarzenia dla *t* D_t	Przeżycie (0 zdarzeń) *t* $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-*t*) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla *t* R_t	Zdarzenia dla *t* D_t	Przeżycie (0 zdarzeń) *t* $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-*t*) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t-1) \cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t - D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854
82	16	1	15/16	0.854·15/16=0.801

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla *t* R_t	Zdarzenia dla *t* D_t	Przeżycie (0 zdarzeń) *t* $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-*t*) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854
82	16	1	15/16	0.854·15/16=0.801
…	15	0	15/15	0.801·15/15=0.801

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854
82	16	1	15/16	0.854·15/16=0.801
…	15	0	15/15	0.801·15/15=0.801
92	15	1	14/15	0.801·14/15=0.748

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854
82	16	1	15/16	0.854·15/16=0.801
…	15	0	15/15	0.801·15/15=0.801
92	15	1	14/15	0.801·14/15=0.748
…	14	0	14/14	0.748·14/14=0.748

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

- $S(t)$ zmienia się tylko dla czasów zdarzeń
- Obserw. cenzurowane zmieniają liczebność zbioru ryzyka
- $S(t)$ nieokreślone dla t>120 (brak obserwacji)
  - 0 jeśli zdarzenie dla największego czasu

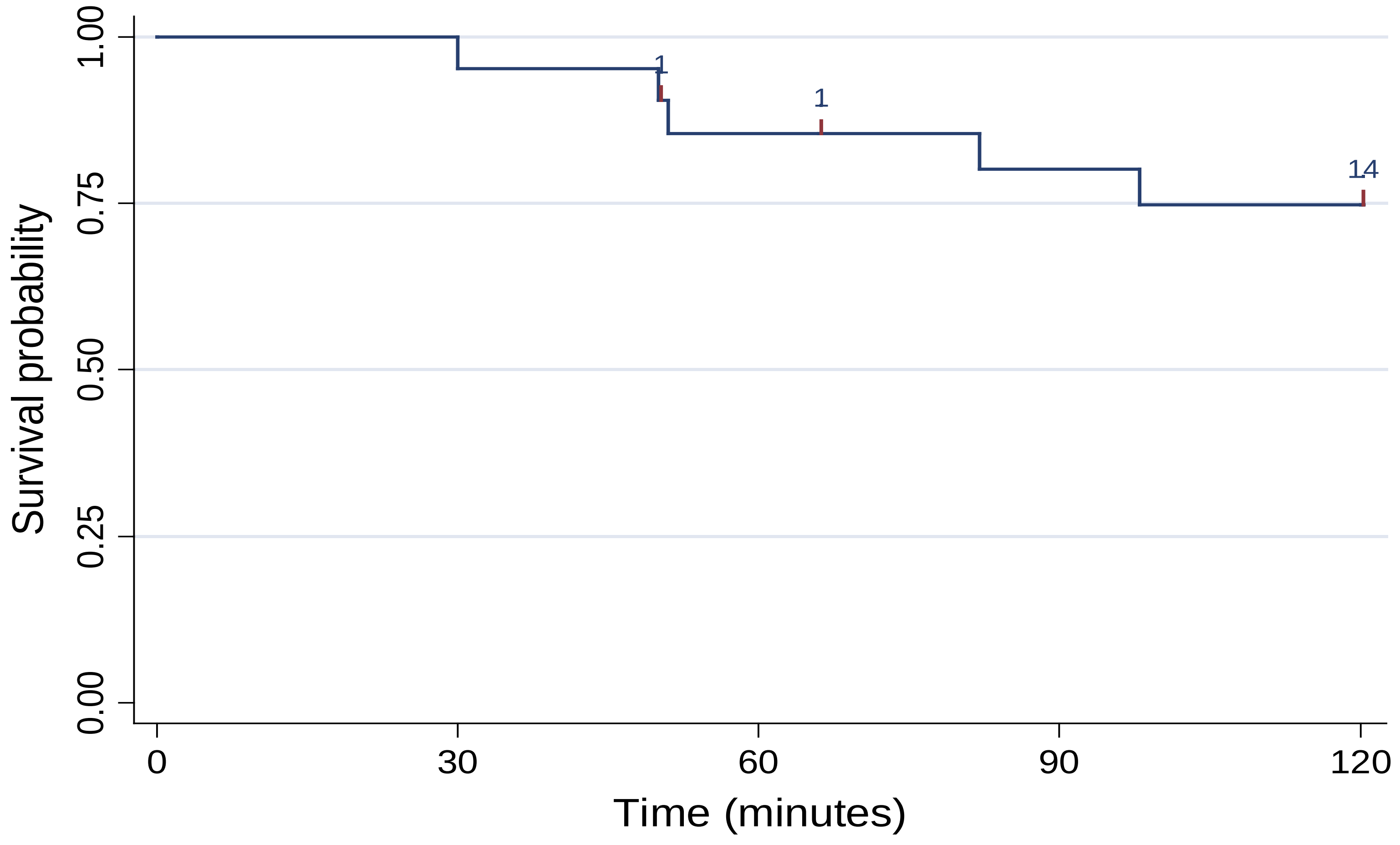
Czas t	Zbiór ryzyka dla t R_t	Zdarzenia dla t D_t	Przeżycie (0 zdarzeń) t $P_t=(R_t-D_t)/R_t$	Prob. przeżycia (brak zdarzeń dla 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
1	21	0	21/21	1·21/21=1
2	21	0	21/21	1·21/21=1
…	21	0	21/21	1·21/21=1
29	21	0	21/21	1·21/21=1
30	21	1	20/21	1·20/21=0.952
31	20	0	20/20	0.952·20/20=0.952
…	20	0	20/20	0.952·20/20=0.952
49	20	0	20/20	0.952·20/20=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
…	17	0	17/17	0.854·17/17=0.854
66	17	0	17/17	0.854·17/17=0.854
67	16	0	16/16	0.854·16/16=0.854
…	16	0	16/16	0.854·16/16=0.854
82	16	1	15/16	0.854·15/16=0.801
…	15	0	15/15	0.801·15/15=0.801
92	15	1	14/15	0.801·14/15=0.748
…	14	0	14/14	0.748·14/14=0.748
120	14	0	14/14	0.748·14/14=0.748

Przykład: choroba lokomocyjna (2)

Osoba	Czas (minuty)	Zdarzenie (1=tak)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

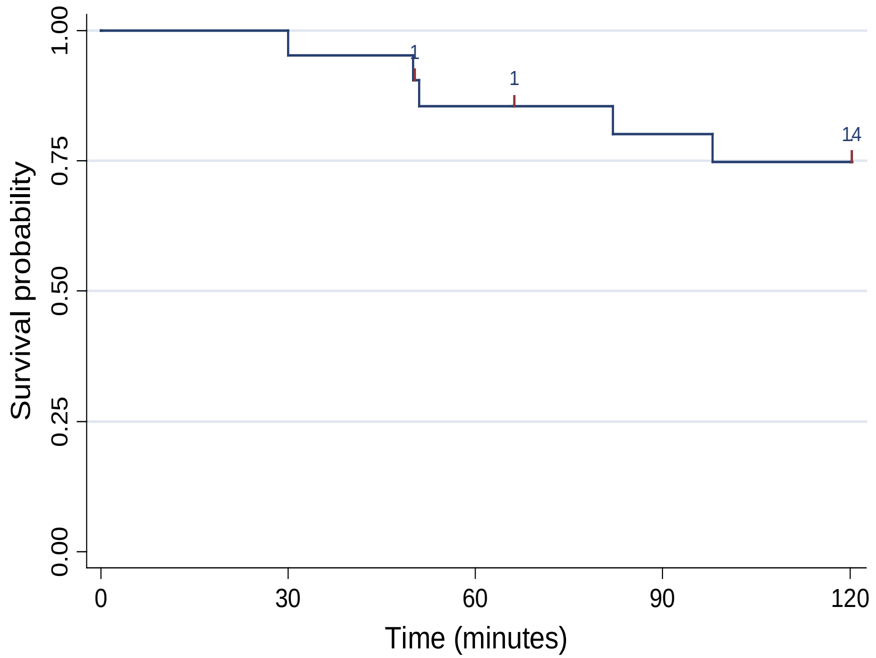
Czas $t_{(j)}$	Zbiór ryzyka dla $t_{(j)}$ n_j	Zdarzenia dla $t_{(j)}$ d_j	Przeżycie $t_{(j)}$ $(n_j - d_j)/n_j$	P-stwo przeżycia $S(t_{(j)})=S(t_{(j-1)})\cdot(n_j - d_j)/n_j$
30	21	1	20/21	1.000·20/21=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
82	16	1	15/16	0.854·15/16=0.801
92	15	1	14/15	0.801·14/15=0.748

Funkcja przeżycia: choroba lokomocyjna



Funkcja przeżycia

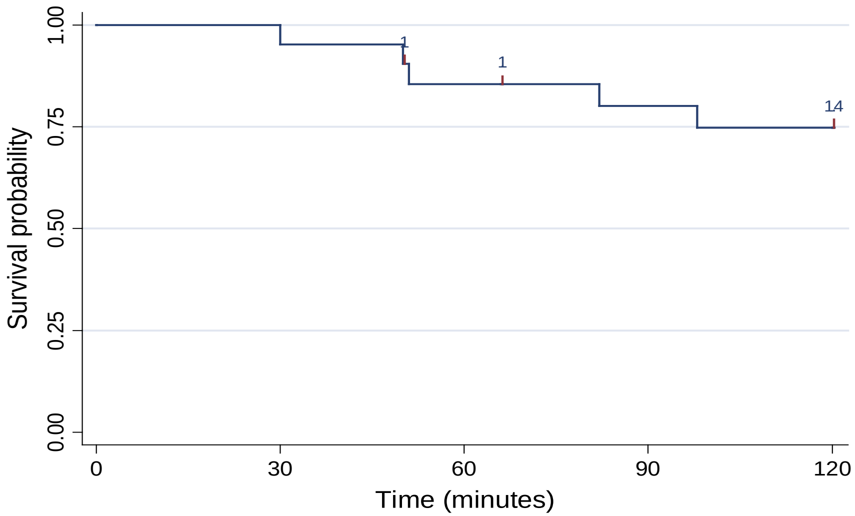
- Wykres nazywany *krzywą przeżycia*
- Charakterystyczne „schodki” dla zaobserwowanych czasów zdarzeń
  - „Wygładzanie” (np.łączenie linią) prowadzi do błędów!



- Uwaga z interpretacją ostatniej części krzywej przeżycia: szacowanie na małej liczbie obserwacji. W szczególności:
  - jeśli ostatnia obserwacja cenzurowana, plateau;
  - jeśli ostatnia obserwacja to zdarzenie, krzywa „spada” do 0.

Przykład: choroba lokomocyjna (3)

Czas $t_{(j)}$	Zbiór ryzyka dla $t_{(j)}$ n_j	Zdarzenia dla $t_{(j)}$ d_j	Przeżycie $t_{(j)}$ $(n_j - d_j)/n_j$	P-stwo przeżycia $S(t_{(j)})=S(t_{(j-1)})\cdot(n_j - d_j)/n_j$
30	21	1	20/21	1.000·20/21=0.952
50	20	1	19/20	0.952·19/20=0.905
51	18	1	17/18	0.905·17/18=0.854
82	16	1	15/16	0.854·15/16=0.801
92	15	1	14/15	0.801·14/15=0.748



- Oszacowanie p-stwa braku torsji przez dłużej niż 30 minut: 95.2%
- Dłużej niż 60 minut: 85.4%
- Dłużej niż 90 minut: 80.1%

Czas przeżycia w próbie klinicznej (cd.)

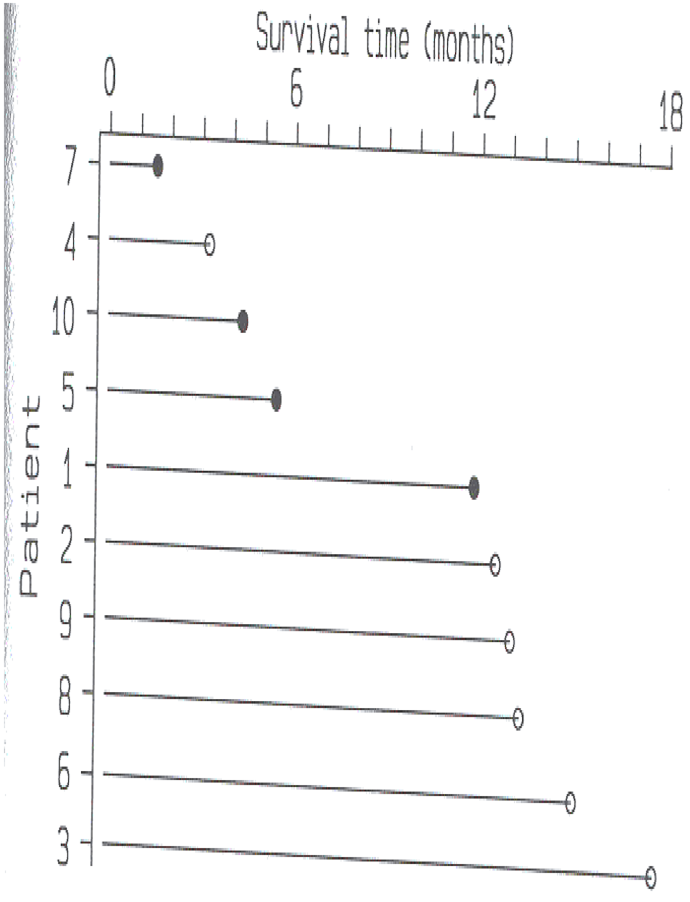


Figure 13.2 Figure 13.1 reorganized to correspond to method of analysis.

Pacjent j	Czas przeżycia (mies.) t_j	Wsk. zgonu (1=tak) δ_j
1	11.8	1
2	12.5	0
3	17.6	0
4	3.2	0
5	5.4	1
6	15.0	0
7	1.5	1
8	13.3	0
9	13.0	0
10	4.3	1

Czas przeżycia w próbie klinicznej (cd.)

Czas (mies.)	Zgon (1=tak)	S(t)
1.5	1	90.0%
3.2	0	90.0%
4.3	1	78.7%
5.4	1	67.5%
11.8	1	56.2%
12.5	0	56.2%
13.0	0	56.2%
13.3	0	56.2%
15.0	0	56.2%
17.6	0	56.2%

- 10 pacjentów, 4 zgony

- „Naiwne" oszacowania p-stwa przeżycia 6 miesięcy:

(3.2 = zgon) $\rightarrow$ *6/10 = 60%*

(3.2 = 6 mies.) $\rightarrow$ *7/10 = 70%*

(3.2 zignorowane) $\rightarrow$ *6/9 = 66.7%*

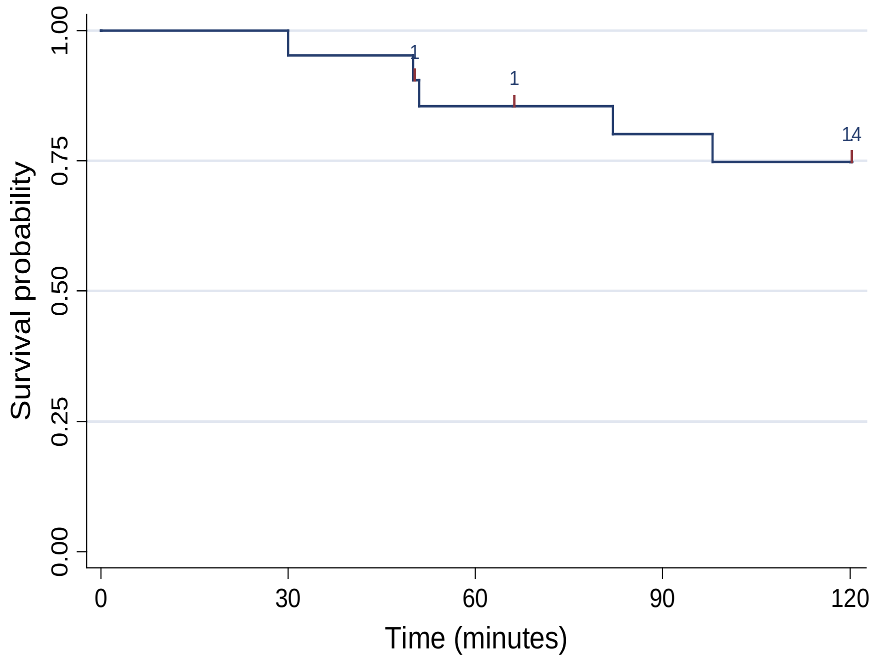
- Kaplan-Meier: 67.5%

Mediana czasu do wystąpienia zdarzenia

- Często używana
- $S(t_{med}) = 50\%$
- Oszacowanie: najkrótszy zaobserwowany czas zdarzenia t, dla którego $S(t) < 50\%$
  - Jeśli $S(t)$=50% dla $[t_{(k)}, t_{(k+1)}]$, to $t_{med} = \{ t_{(k)} + t_{(k+1)} \} / 2$
- Problemy:
  - brak estymatora jeśli $S(t) > 50\%$ dla wszystkich t
  - błąd standardowy (SE) trudny do oszacowania
  - bardzo duża zmienność oszacowania

Mediana czasu do wystąpienia zdarzenia: choroba lokomocvina

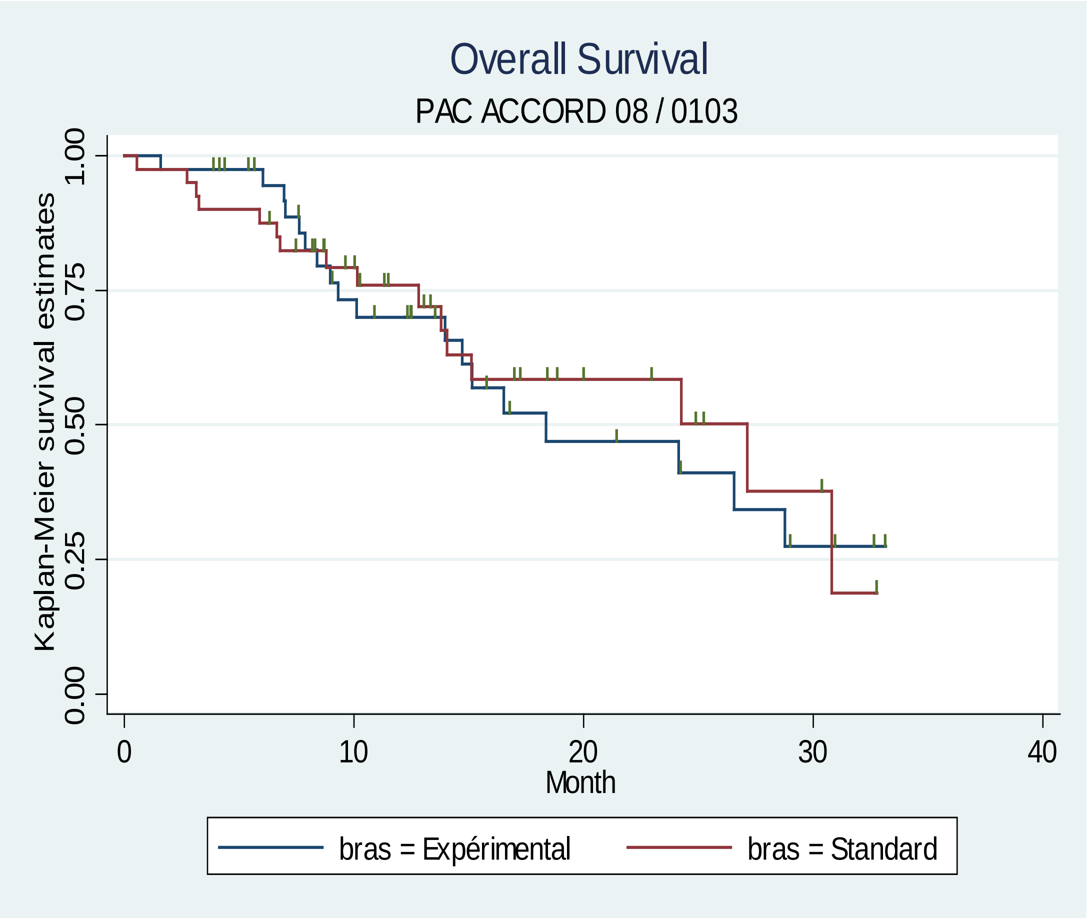
- Brak oszacowania



Mediana czasu do wystąpienia zdarzenia: przykład

- Mediana dla grupy badanej: około 18 mies.
- Mediana dla grupy kontrolnej: około 27 mies.

- ALE
jeśli którakolwiek z obserwacji cenzurowanych <27 mies. w grupie kontrolnej będzie zdarzeniem, mediana znacznie się skróci !



Błąd standardowy estymatora prawdopodobieństwa przeżycia (2)

- Najprostsze oszacowanie przy użyciu przybliżenia z rozkładu dwumianowego:

$$\mathrm{S\hat{E}}\left\{\hat{S}(t_{(j)})\right\} = \sqrt{\hat{S}(t_{(j)})\left\{1 - \hat{S}(t_{(j)})\right\}/n_j}$$

Błąd dwumianowy: choroba lokomocyjna

Czas $t_{(j)}$	Zbiór ryzyka dla $t_{(j)}$ n_j	P-stwo przeżycia $S(t)$	SE (dwumianowy)
30	21	0.9524	$\{ 0.9524(1-0.9524) / 21 \}^{1/2} = 0.0464$
50	20	0.9048	$\{ 0.9048(1-0.9048) / 20 \}^{1/2} = 0.0656$
51	18	0.8545	$\{ 0.8545(1-0.8545) / 18 \}^{1/2} = 0.0831$
82	16	0.8011	$\{ 0.8011(1-0.8011) / 16 \}^{1/2} = 0.0998$
92	15	0.7477	$\{ 0.7477(1-0.7477) / 15 \}^{1/2} = 0.1121$

- Błąd standardowy wzrasta ze wzrostem czasu
  - mniejsza precyzja oszacowania wynikająca z mniejszego zbioru ryzyka

Wzór Greenwooda

$$\hat{S}(t) = \prod_{t_{(j)} \le t} \left(\frac{n_j - d_j}{n_j} \right) \equiv \prod_{t_{(j)} \le t} \hat{\pi}_j \quad \Rightarrow \quad \mathrm{Var}\left\{ \ln \hat{S}(t) \right\} \approx \sum_{t_{(j)} \le t} \mathrm{Var}\left(\ln \hat{\pi}_j \right)$$

- Metoda „delta”:

$$\hat{\mathrm{V}}\mathrm{ar}\left(\ln \hat{\pi}_j \right) = \hat{\mathrm{V}}\mathrm{ar}\left(\hat{\pi}_j \right) / \hat{\pi}_j^2 = \left\{ \hat{\pi}_j \left(1 - \hat{\pi}_j \right) / n_j \right\} / \hat{\pi}_j^2$$

$$\Downarrow$$

$$\hat{\mathrm{V}}\mathrm{ar}\left\{ \ln \hat{S}(t) \right\} = \sum_{t_{(j)} \le t} \frac{d_j}{n_j \left(n_j - d_j \right)}$$

- Metoda „delta”:

$$\hat{\mathrm{V}}\mathrm{ar}\left\{ \hat{S}(t) \right\} = \left\{ \hat{S}(t) \right\}^2 \hat{\mathrm{V}}\mathrm{ar}\left\{ \ln \hat{S}(t) \right\} = \left\{ \hat{S}(t) \right\}^2 \sum_{t_{(j)} \le t} \frac{d_j}{n_j \left(n_j - d_j \right)}$$

Wzór Greenwooda: choroba lokomocyjna

$t_{(j)}$	n_j	d_j	$n_j - d_j$	$S(t)$	$d_j/\{n_j(n_j - d_j)\}$	$[\sum_j d_j/\{n_j(n_j - d_j)\}]^{½}$	SE
30	21	1	20	0.9524	0.0024	0.0488	0.9524·0.0488=0.0465
50	20	1	19	0.9048	0.0026	0.0707	0.9048·0.0707=0.0640
51	18	1	17	0.8545	0.0033	0.0910	0.8545·0.0910=0.0778
82	16	1	15	0.8011	0.0042	0.1116	0.8011·0.1116=0.0894
92	15	1	14	0.7477	0.0048	0.1312	0.7477·0.1312=0.0981

Własności estymatora Kaplana-Meiera

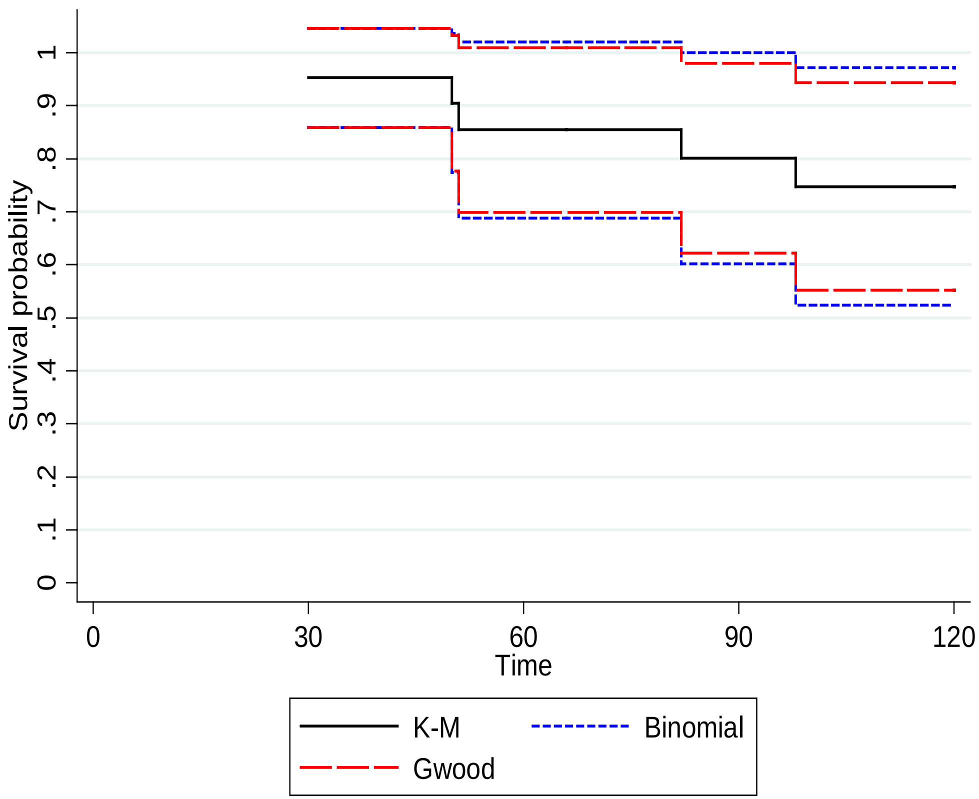
- Estymator nieparametryczny oparty na maksimum funkcji wiarogodności (MLE)
- Wymaga założenia niezależności obserwacji i niezależnego i nie-informatywnego cenzurowania
- Dla „dużych *n*”,

$$\left\{\hat{S}(t) - S(t)\right\} / \sqrt{\mathrm{Var}\left\{\hat{S}(t)\right\}} \xrightarrow{D} N(0,1)$$

  - dowód przy użyciu teorii martyngałów
  - SE z wzoru Greenwooda
  - 95% przedział ufności $\hat{S}(t) \pm 1.96\sqrt{\mathrm{Var}\left\{\hat{S}(t)\right\}}$

Błąd dwumianowy vs. Greenwood: choroba lokomocyjna

Czas $t_{(j)}$	Zbiór ryzyka n_j	P-stwo przeżycia $S(t)$	SE (dwum)	SE (G'wood)
30	21	0.952	0.046	0.047
50	20	0.905	0.066	0.064
51	18	0.854	0.083	0.078
82	16	0.801	0.100	0.089
92	15	0.748	0.112	0.098



- 95% przedział ufności:

$$S(t) \pm 1.96 \cdot \mathrm{SE}$$

- Przybliżenie rozkładem dwumianowym daje szersze granice
- Problem: górna granica > 1

Błąd standardowy estymatora prawdopodobieństwa przeżycia (2)

- Mamy $\hat{\mathrm{V}}\mathrm{ar}\left\{\ln \hat{S}(t)\right\} = \sum_{t_{(j)} \le t} \frac{d_j}{n_j\left(n_j - d_j\right)}$

- Stąd 95% przedział ufności:

$$\ln \hat{S}(t) \pm 1.96 \sqrt{\sum_{t_{(j)} \le t} \frac{d_j}{n_j\left(n_j - d_j\right)}}$$

- Po przekształceniu, przedział dla S(t):

$$\hat{S}(t) e^{\pm 1.96 \cdot \sqrt{\mathrm{Var}\left[\ln \hat{S}(t)\right]}}$$

Błąd standardowy estymatora prawdopodobieństwa przeżycia (3)

- Metoda „delta”:

$$\hat{\mathrm{V}}\mathrm{ar}\left[\ln\left\{-\ln \hat{S}(t)\right\}\right]=\left\{\ln \hat{S}(t)\right\}^{-2}\hat{\mathrm{V}}\mathrm{ar}\left\{\ln \hat{S}(t)\right\}$$

- Otrzymujemy 95% przedział ufności:

$$\ln\left\{-\ln \hat{S}(t)\right\}\pm 1.96\sqrt{\left\{\ln \hat{S}(t)\right\}^{-2}\sum_{t_{(j)}\leq t}\frac{d_j}{n_j\left(n_j-d_j\right)}}$$

- Po przekształceniu, przedział dla S(t):

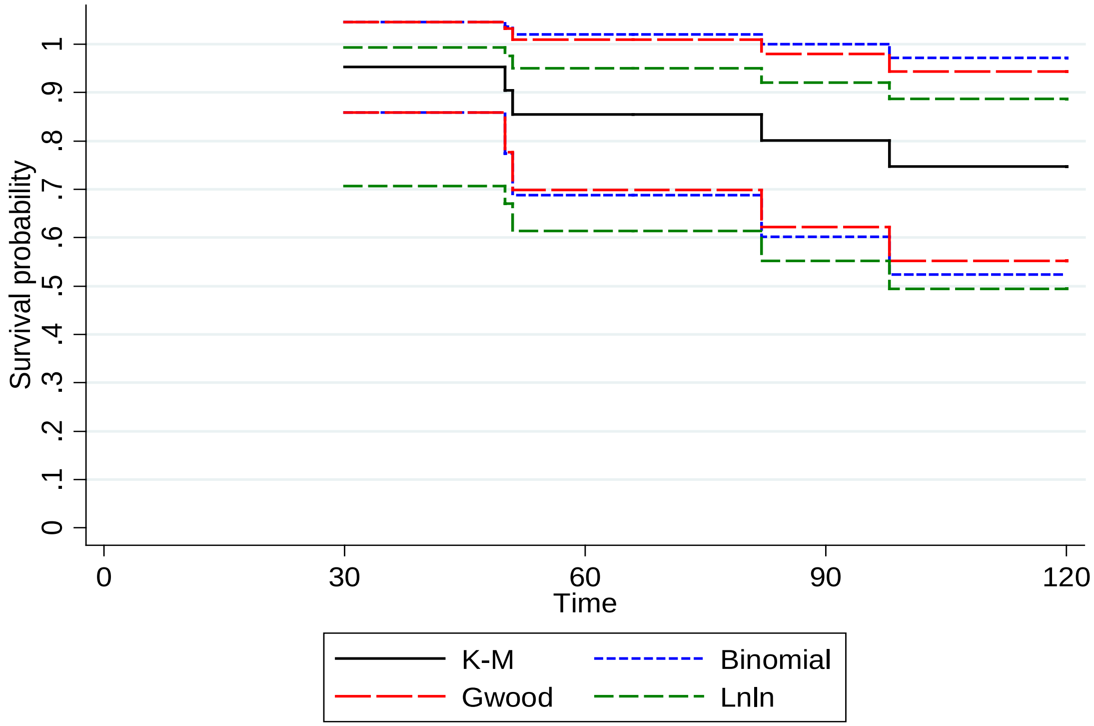
$$\hat{S}(t)^{\exp\left(\pm 1.96\cdot\sqrt{\mathrm{Var}\left[\ln\left\{-\ln \hat{S}(t)\right\}\right]}\right)}$$

Błąd standardowy log(-log *S*): choroba lokomocyjna

$t_{(j)}$	n_j	d_j	$d_j/\{n_j(n_j - d_j)\}$	$[\sum_j d_j/\{n_j(n_j - d_j)\}]^{½}$	$S(t)$	SE
30	21	1	0.0024	0.0488	0.9524	0.0488 / | ln(0.9524) |=1.000
50	20	1	0.0026	0.0707	0.9048	0.0707 / | ln(0.9048) |=0.707
51	18	1	0.0033	0.0910	0.8545	0.0910 / | ln(0.8545) |=0.579
82	16	1	0.0042	0.1116	0.8011	0.1116 / | ln(0.8011) |=0.503
92	15	1	0.0048	0.1312	0.7477	0.1312 / | ln(0.9524) |=0.451

95% przedział ufności dla $S(t)$: choroba lokomocyjna

t	$S(t)$	95% CI (dwum)	95% CI (G'wood)	95% CI (Lnln)
30	0.952	[0.861, 1.043]	[0.861, 1.043]	[0.707, 0.993]
50	0.905	[0.776, 1.033]	[0.779, 1.030]	[0.670, 0.975]
51	0.854	[0.692, 1.017]	[0.702, 1.007]	[0.613, 0.951]
82	0.801	[0.605, 0.997]	[0.626, 0.976]	[0.552, 0.921]
92	0.748	[0.528, 0.967]	[0.555, 0.940]	[0.495, 0.887]



95% przedział ufności dla *S*(*t*): choroba lokomocyjna (cd.)

t	*S(t)*	95% CI (dwum)	95% CI (G'wood)	95% CI (Lnln)
30	0.952	[0.861, 1.043]	[0.861, 1.043]	[0.707, 0.993]
50	0.905	[0.776, 1.033]	[0.779, 1.030]	[0.670, 0.975]
51	0.854	[0.692, 1.017]	[0.702, 1.007]	[0.613, 0.951]
82	0.801	[0.605, 0.997]	[0.626, 0.976]	[0.552, 0.921]
92	0.748	[0.528, 0.967]	[0.555, 0.940]	[0.495, 0.887]

- UWAGA: *punktowe* przedziały ufności
  - 95% przedziały ufności dla pojedyńczego czasu, nie dla całej krzywej przeżycia
  - łączenie górnych/dolnych granic nie daje poprawnego obszaru ufności dla funkcji przeżycia

95% obszar ufności dla $S(t)$

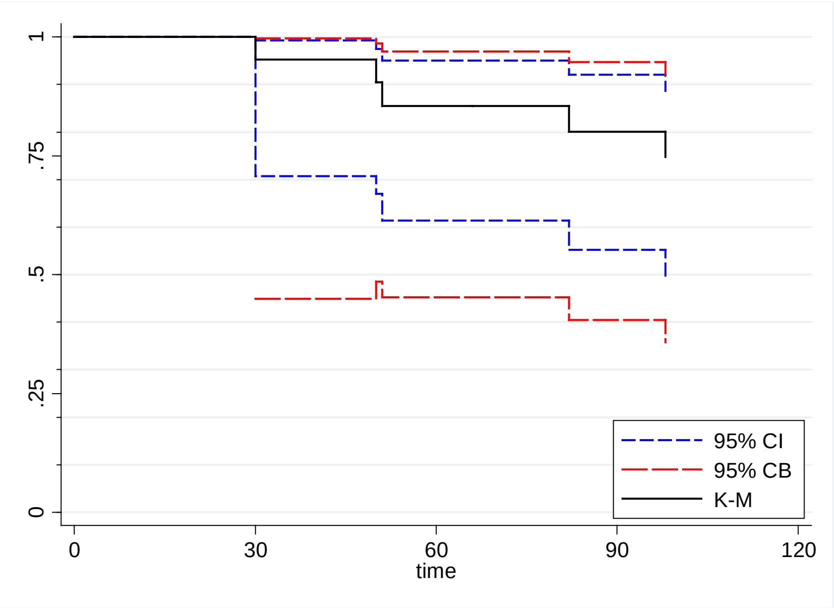
- Obszar, który w 95% przypadków zawiera funkcję przeżycia dla określonego przedziału czasu $[t_L, t_U]$:

$$P\{L(t) \leq S(t) \leq U(t), \text{ dla } t_L \leq t \leq t_U\} = 95\%$$

- Konstrukcje z wykorzystaniem procesów stochastycznych
  - most Browna

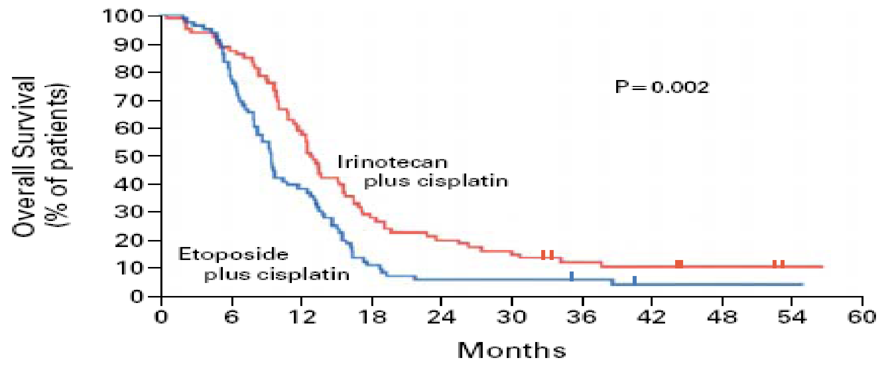
95% obszar ufności dla *S*(*t*): choroba lokomocyjna

t	*S(t)*	95% CI (Lnln)	95% CB (Lnln)
30	0.952	[0.707, 0.993]	[0.449, 0.997]
50	0.905	[0.670, 0.975]	[0.485, 0.986]
51	0.854	[0.613, 0.951]	[0.452, 0.969]
82	0.801	[0.552, 0.921]	[0.404, 0.947]
92	0.748	[0.495, 0.887]	[0.358, 0.921]



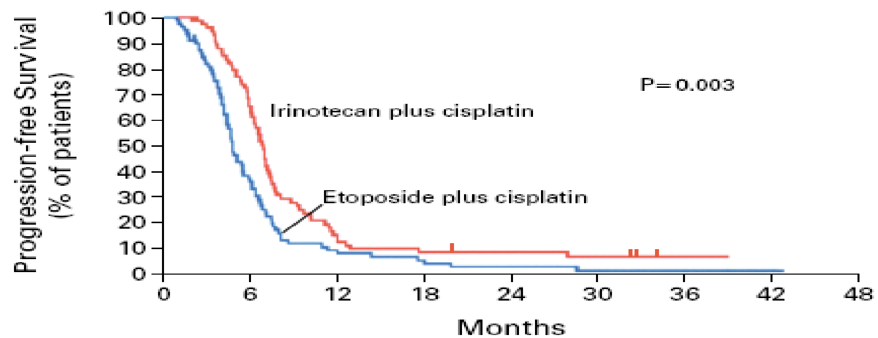
Próba kliniczna SCLC

Noda *et al.*, *NEJM* (2002)



No. at Risk							
Irinotecan plus cisplatin	77	67	45	21	15	11	7
Etoposide plus cisplatin	77	60	29	8	4	4	3

Figure 1. Overall Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.



No. at Risk				
Irinotecan plus cisplatin	77	47	9	9
Etoposide plus cisplatin	77	27	6	2

Figure 2. Progression-free Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.

Parametryczne szacowanie $S(t)$

- Jeśli zakładamy postać rozkładu p-stwa czasu

- Przykłady:
  - $T^* \sim$ wykładniczy: $S(t) = exp(-\lambda t)$
  - $T^* \sim$ Weibull: $S(t) = exp\{ -(\lambda t)^p \}$

- By oszacować $S(t)$, musimy oszacować parametry

Szacowanie funkcji przeżycia dla rozkładu wykładniczego (1)

- Jeśli T^* ma rozkład wykładniczy, to $S(t) = exp(-\lambda t)$
  - Założenie: *współczynnik intensywności* λ zdarzeń <u>stały</u> w czasie

- (Częściowa) Funkcja wiarogodności:

$$L(\lambda) = \prod_{j=1}^{n} f\left(t_j\right)^{\delta_j} S\left(t_j\right)^{1-\delta_j} = \prod_{j=1}^{n} \lambda\left(t_j\right)^{\delta_j} S\left(t_j\right) = \prod_{j=1}^{n} \lambda^{\delta_j} e^{-\lambda t_j}$$

Szacowanie funkcji przeżycia dla rozkładu wykładniczego (2)

- Logarytm (częściowej) funkcji wiarogodności:

$$l(\lambda) \equiv \ln L(\lambda) = \sum_{j=1}^{n}\left(\delta_j \ln\lambda - \lambda t_j\right) = (\ln\lambda)\sum_{j=1}^{n}\delta_j - \lambda\sum_{j=1}^{n} t_j$$

- Maksimum dla:

$$\frac{dl(\lambda)}{d\lambda} = 0 \Leftrightarrow \hat{\lambda}^{-1}\sum_{j=1}^{n}\delta_j - \sum_{j=1}^{n} t_j = 0 \Leftrightarrow \hat{\lambda} = \sum_{j=1}^{n}\delta_j \Big/ \sum_{j=1}^{n} t_j$$

- Oszacowanie wariancji:

$$\left[\left\{-\frac{d^2 l(\lambda)}{d\lambda^2}\right\}_{|\lambda=\hat{\lambda}}\right]^{-1} = \left\{\sum_{j=1}^{n}\delta_j / \hat{\lambda}^2\right\}^{-1} = \sum_{j=1}^{n}\delta_j \Big/ \left(\sum_{j=1}^{n} t_j\right)^2$$

Szacowanie funkcji przeżycia dla rozkładu wykładniczego (3)

- Proste oszacowanie λ:

  $\sum\delta_j / \sum t_j$ = *liczba zdarzeń / sumaryczny czas*

- Estymator błędu standardowego: $(\sum \delta_j)^{½} / \sum t_j$

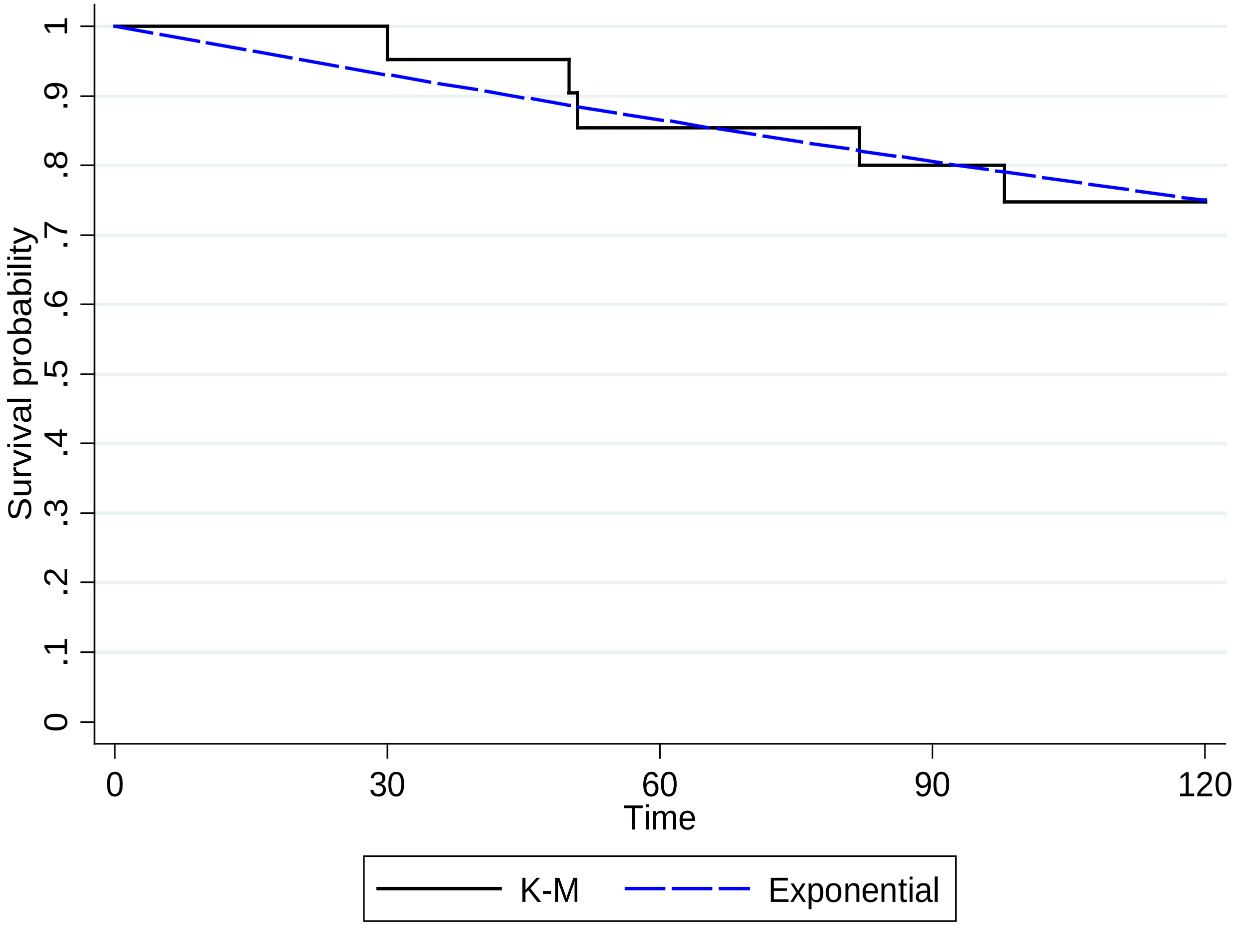
Rozkład wykładniczy: choroba lokomocyjna

- $\sum\delta_j / \sum t_j = 5/2101 = 0.0024$
- SE = $5^{½}$ / 2101 = 0.0011
- Funkcja przeżycia $S(t) = e^{-0.0024t}$

j	t_j	δ_j
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0
Total	2101	5

Czas	K-M S(t)	Exp S(t)
30	0.952	0.931
50	0.905	0.887
51	0.854	0.885
82	0.801	0.821
92	0.748	0.750

Rozkład wykładniczy: choroba lokomocyjna (cd.)



Szacowanie funkcji hazardu

- Załóżmy stałą funkcję hazardu dla $[t_{(j)}, t_{(j+1)})$.
- Z rozkładu wykładniczego

$$\hat{\lambda}_j = \frac{d_j}{n_j (t_{(j+1)} - t_{(j)})}$$

  - uwaga: nie działa dla ostatniego czasu zdarzenia!
  - Oszacowanie wariancji z rozkł. dwumianowego:

$$\hat{\mathrm{V}}\mathrm{ar}(\hat{\lambda}_j) = \frac{1}{n^2{}_j (t_{(j+1)} - t_{(j)})^2} \frac{d_j (n_j - d_j)}{n_j} = \hat{\lambda}^2 \frac{n_j - d_j}{n_j d_j}$$

- W praktyce, szacowanie funkcji hazardu jest rzadko stosowane (duża zmienność).

Szacowanie funkcji skumulowanego hazardu

- Ponieważ $\Lambda(t)$= -ln$S(t)$, używając K-M, mamy

$$\hat{\Lambda}_{K-M}(t) = -\sum_{t_{(j)} \leq t} \ln\left(\frac{n_j - d_j}{n_j}\right)$$

- Jako że ln(1+x) ≈ x, dostajemy estymator Nelsona-Aalena

$$\hat{\Lambda}_{N-A}(t) = \sum_{t_{(j)} \leq t} d_j / n_j$$

  - Daje alternatywny, do K-M, estymator $S(t)$: $\hat{S}(t) = e^{-\hat{\Lambda}_{N-A}(t)}$

- Z różnic dla estymatorów $\Lambda(t)$ dla $[t_{(j)}, t_{(j+1)})$ możemy uzyskać oszacowanie funkcji hazardu.

Własności estymatorów funkcji skumulowanego hazardu

- Dla estymatora Nelsona-Aalena $\hat{\mathrm{V}}\mathrm{ar}\,\hat{\Lambda}_{\mathrm{N-A}}(t) = \sum_{t_{(j)} \le t} \frac{d_j (n_j - d_j)}{n_j^3}$
  - Alternatywne oszacowanie:

$$\hat{\mathrm{V}}\mathrm{ar}\,\hat{\Lambda}_{\mathrm{N-A}}(t) = \left\{\hat{S}(t)\right\}^2 \sum_{t_{(j)} \le t} \frac{d_j}{n_j^2}$$

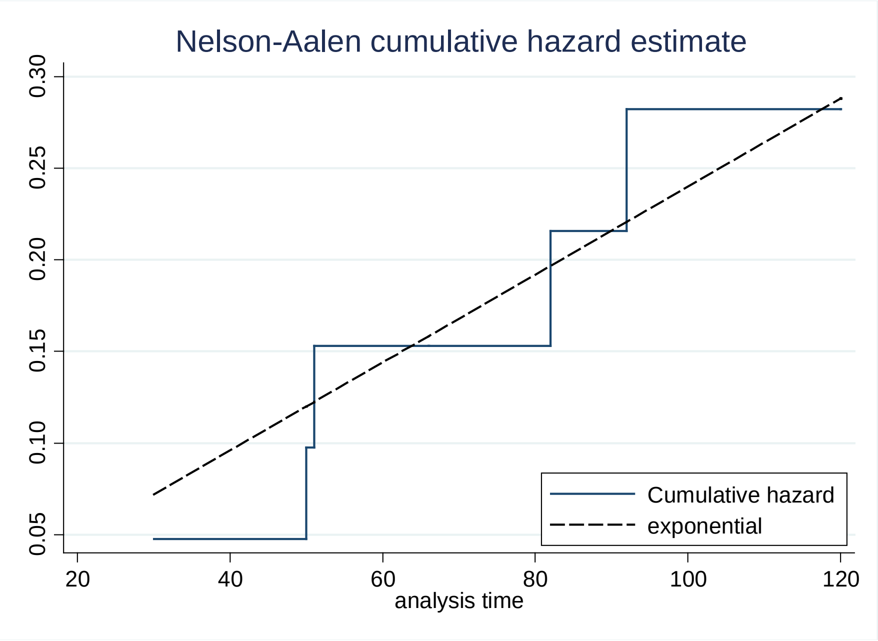
- Można pokazać, że $\hat{\Lambda}(t) \xrightarrow{P} \Lambda(t)$

oraz że

$$\left\{\hat{\Lambda}(t) - \Lambda(t)\right\} / \sqrt{\mathrm{Var}\left\{\hat{\Lambda}(t)\right\}} \xrightarrow{D} N(0,1)$$

Skumulowana funkcja hazardu: choroba lokomocyjna

Time	Nelson-Aalen Cum. Haz.	Std. Error	[95% Conf.	Int.]
30	0.0476	0.0476	0.0067	0.3381
50	0.0976	0.0690	0.0244	0.3905
51	0.1532	0.0886	0.0493	0.4761
66	0.1532	0.0886	0.0493	0.4761
82	0.2157	0.1084	0.0805	0.5778
92	0.2823	0.1273	0.1167	0.6832
120	0.2823	0.1273	0.1167	0.6832



Metoda „tablicy życia”

- Stosowana do szacowania *S(t)* dla danych zgrupowanych
  - Gdy nie mamy dokładnych czasów, lecz jedynie informację o liczbie zdarzeń dla przedziałów czasu.

Metoda „tablicy życia”: przykład (1)

- Hipotetyczne dane dla 300 osób po przeszczepie serca

- Problem: niektóre osoby nie były obserwowane przez cały przedział czasu (*wycofania*)

Przedział *[t, t+2)*	*Pod obs. w chwili t* R_t	*Zmarli w* *[t, t+2)* D_t	*Wycofani w* *[t, t+2)* W_t
0-2	300	193	6
2-4	101	25	8
4-6	68	12	10
6-8	46	8	10
8-10	28	4	10
10-12	14	3	10
12+	1	0	1

Metoda „tablicy życia”: przykład (2)

- Szacujemy $S(t)$
- Potrzebujemy oszacowanie p-stwa przeżycia przedziału [$t,t+2$)
- Idea: użyjmy $(R_t - D_t) / R_t$
  - Ale są *wycofania*, dla których nie mamy informacji o całym przedziale
  - R_t musi być zmniejszone

Przedział [t, t+2)	*Pod obs. w chwili t* R_t	*Zmarli w [t, t+2)* D_t	*Wycofani w [t, t+2)* W_t
0-2	300	193	6
2-4	101	25	8
4-6	68	12	10
6-8	46	8	10
8-10	28	4	10
10-12	14	3	10
12+	1	0	1

Metoda „tablicy życia”: przykład (3)

- Zakładamy, że wycofani (W_t) są „podobni” do pozostających pod obserwacją.
- I że czasy wycofań są rozłożone jednostajnie w $[t,t+t_0)$.
- Średnio wycofani dodają informację o połowie t_0 do zbioru ryzyka.
- Używamy więc „poprawionych” liczebności zbioru ryzyka:

$$R'_t = R_t - W_t / 2$$

t	$[t, t+2)$	R_t	D_t	W_t	R'_t
0	0-2	300	193	6	297
2	2-4	101	25	8	97
4	4-6	68	12	10	63
6	6-8	46	8	10	41
8	8-10	28	4	10	23
10	10-12	14	3	10	9
12	12+	1	0	1	0.5

Metoda „tablicy życia”:

- Po poprawce na wycofania, definiujemy:

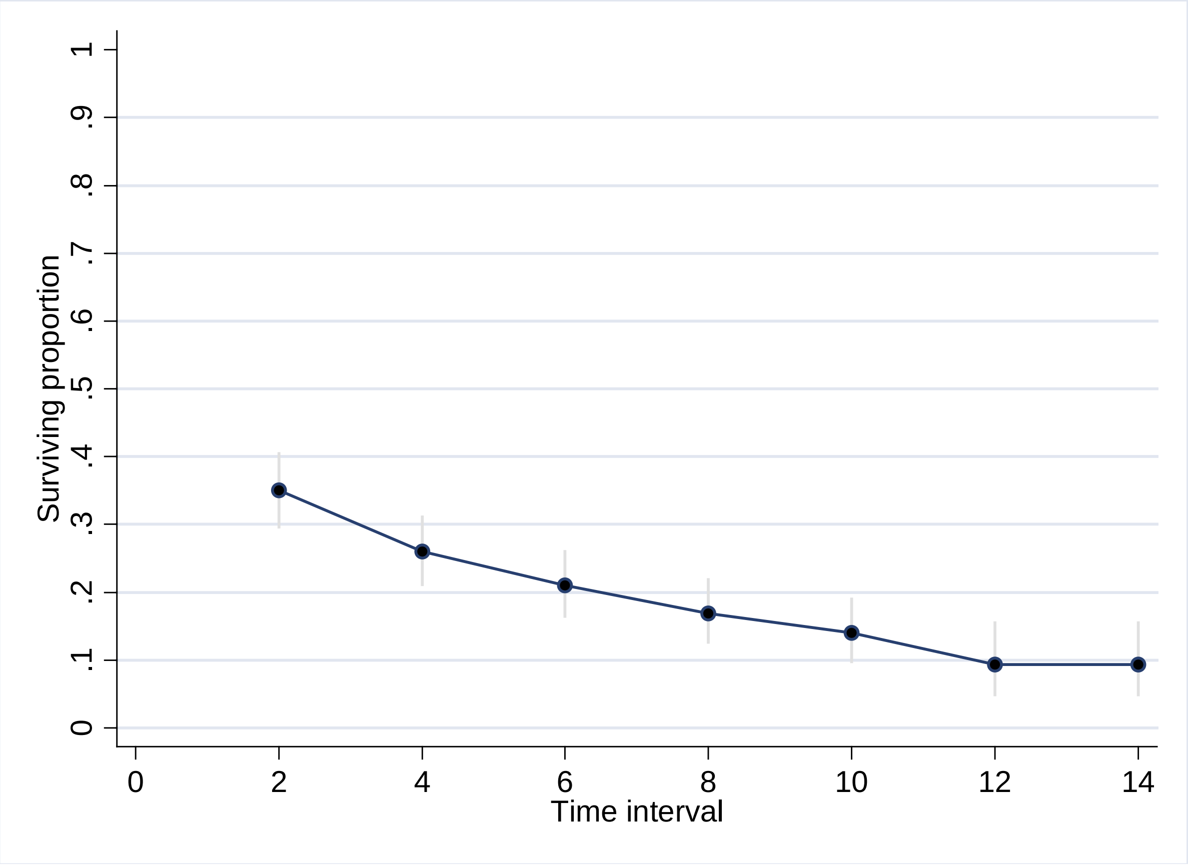
$$P\{\, przeżycie\ [t, t+2)\, \} = P_t = (\, R'_t - D_t \,) \,/\, R'_t$$

- W rezultacie otrzymujemy:

$$S(t) = S(t-2) \cdot P_{t-2} = S(0) \cdot P_0 \cdot \ldots \cdot P_{t-2}$$

t	$[t, t+2)$	R_t	D_t	W_t	R'_t	P_t	$S(t)$
0	0-2	300	193	6	297	104/297=0.35	1
2	2-4	101	25	8	97	72/97=0.74	1 · 0.35=0.35
4	4-6	68	12	10	63	51/63=0.81	0.35 · 0.74=0.26
6	6-8	46	8	10	41	33/41=0.81	0.26 · 0.81=0.21
8	8-10	28	4	10	23	19/23=0.83	0.21 · 0.81=0.17
10	10-12	14	3	10	9	6/9=0.67	0.17 · 0.83=0.14
12	12+	1	0	1	0.5	1	0.14 · 0.67=0.09

Krzywa przeżycia dla metody „tablicy życia”



Błąd standardowy dla oszacowania prawdopodobieństwa przeżycia metodą „tablicy życia”

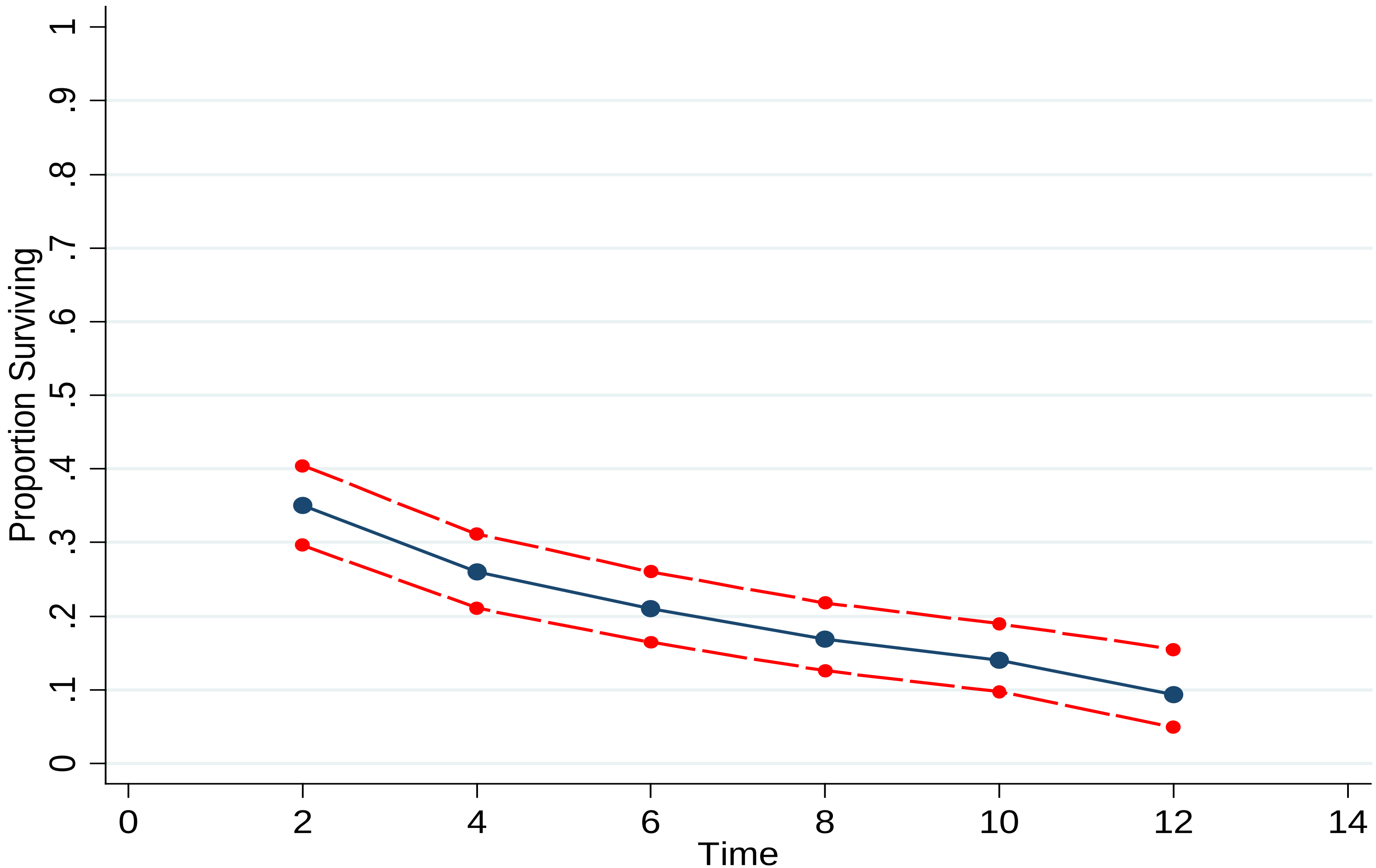
- Używamy „poprawionych” zbiorów ryzyka we wzorze Greenwooda:

$$\hat{\mathrm{V}}\mathrm{ar}\left\{\hat{S}(t)\right\} = \left\{\hat{S}(t)\right\}^2 \sum_{s \le t} \frac{D_s}{R'_s\left(R'_s - D_s\right)}$$

Błąd standardowy dla metody „tablicy życia”: przykład (1)

t	$[t, t+2)$	$S(t)$	D_t	R'_t	$D_t/\{(R'_t - D_t)R'_t\}$	$[\sum D_s/\{R'_t(R'_s - D_s)\}]^{1/2}$	SE
0	0-2	1	193	297	0.0062	0.0790	
2	2-4	0.35	25	97	0.0036	0.0991	0.35 · 0.0790=0.0277
4	4-6	0.26	12	63	0.0037	0.1165	0.26 · 0.0991=0.0258
6	6-8	0.21	8	41	0.0059	0.1396	0.21· 0.1165=0.0245
8	8-10	0.17	4	23	0.0092	0.1692	0.17 · 0.1396=0.0236
10	10-12	0.14	3	9	0.0556	0.2901	0.14 · 0.1692=0.0237
12	12+	0.09					0.09 · 0.2901=0.0271

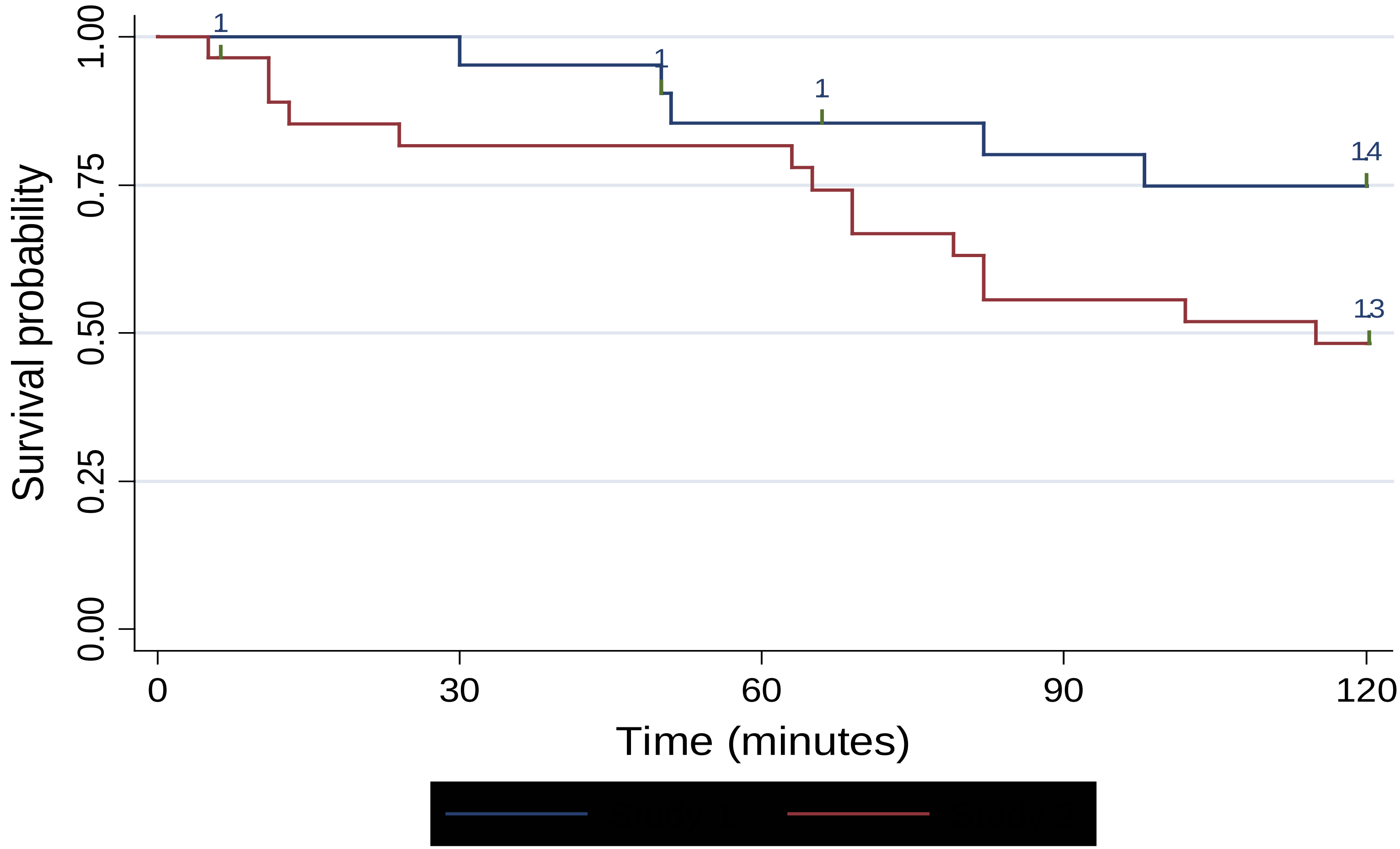
Błąd standardowy dla metody „tablicy życia”: przykład (2)



Choroba lokomocyjna: druga próba

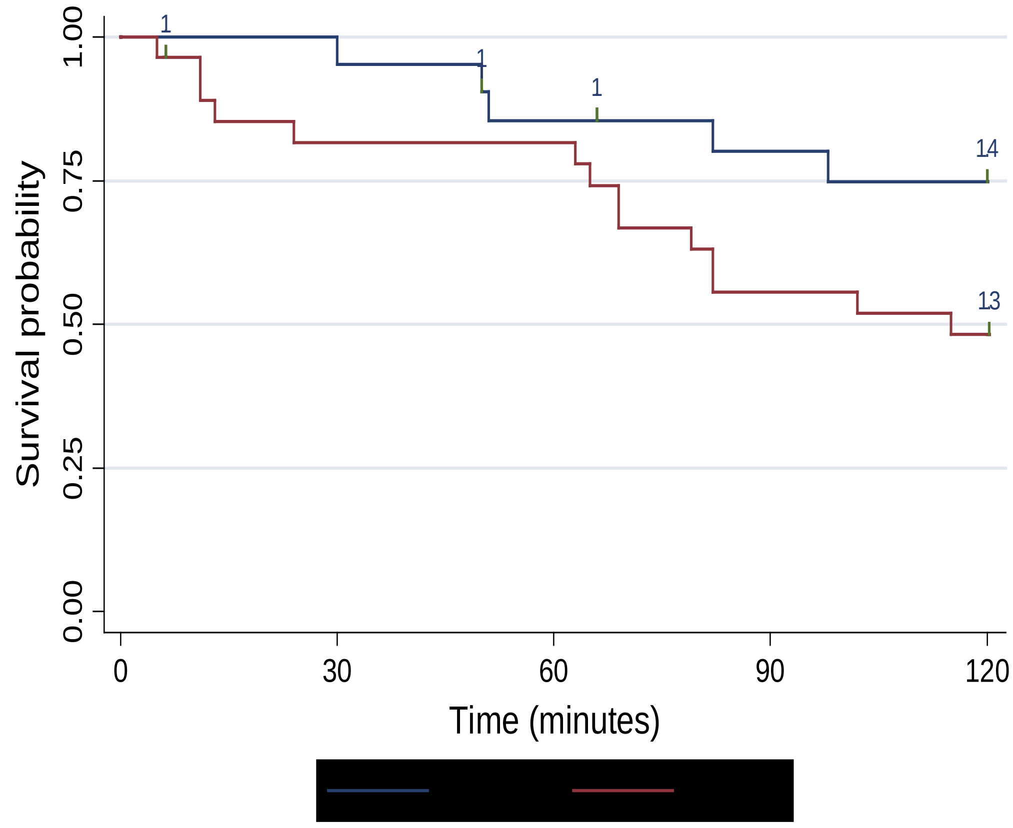
- Doświadczenie z „kołysaniem” powtórzona dla 28 osób…
- … z 2 razy wyższą częstotliwością (0.333 Hz) i przyśpieszeniem (0.222 G) ☹ ☹
- Obserwacje (<u>cenzurowane</u>):
  5, <u>6</u>, 11, 11, 13, 24, 63, 65, 69, 69, 79, 82, 82, 102, 115, <u>120</u> (x13)
- Poprzednie doświadczenie:
  30, 50, <u>50</u>, 51, <u>66</u>, 82, 92, <u>120</u> (x14)
- Czasy w drugim doświadczeniu wydają się krótsze…

Choroba lokomocyjna: krzywe przeżycia



Choroba lokomocyjna: różnica w krzywych przezycia

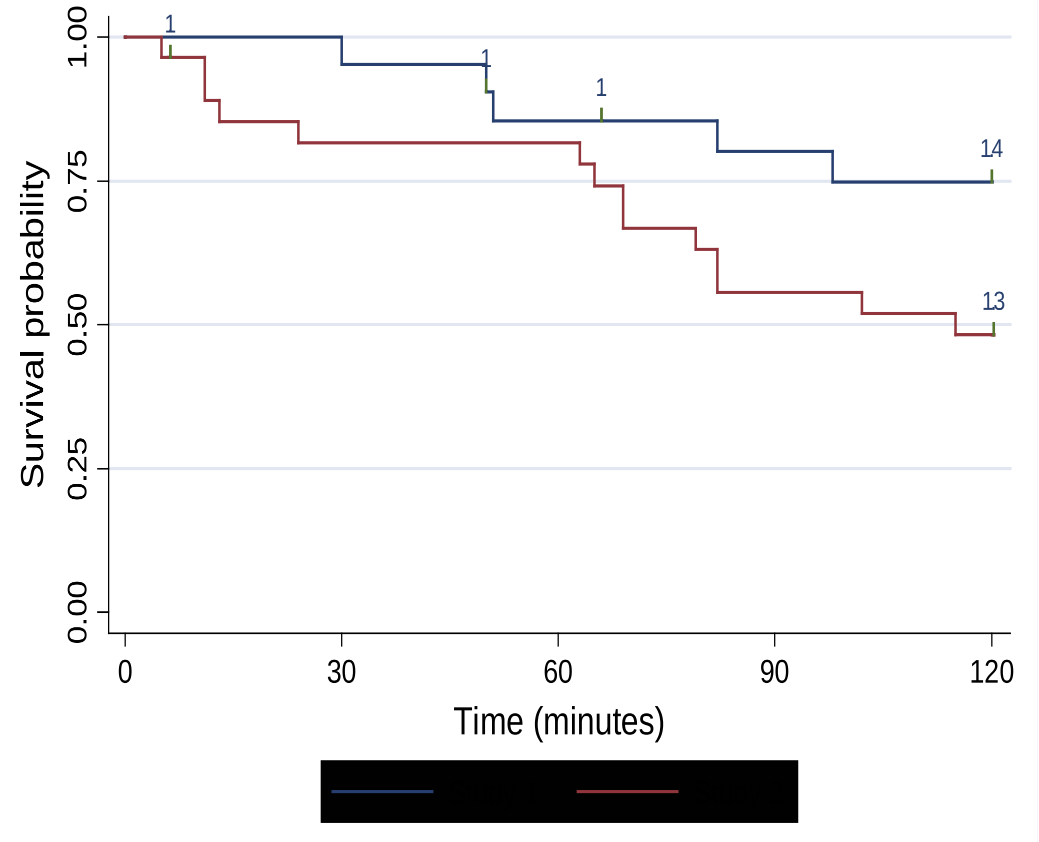
- Dla ustalonego czasu *t* p-stwo “przeżycia” bez torsji było niższe dla drugiego doświadczenia…
- … a więc czas do torsji musiał być średnio krótszy



- 1-sze doświadczenie: $t_{med} > 120$
- 2-gie: $t_{med}=115$

Choroba lokomocyjna: różnica w krzywych przezycia (cd.)

- Ale może różnica jest tylko przypadkowa?

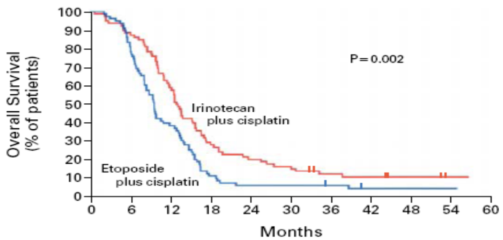


Próba kliniczna SCLC (1)

IRINOTECAN PLUS CISPLATIN COMPARED WITH ETOPOSIDE PLUS CISPLATIN FOR EXTENSIVE SMALL-CELL LUNG CANCER

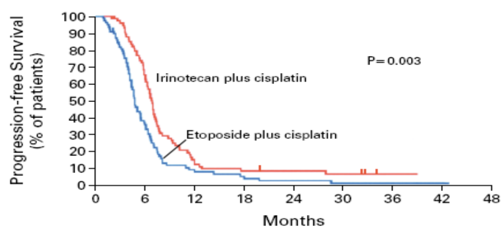
KAZUMASA NODA, M.D., YUTAKA NISHIWAKI, M.D., MASAAKI KAWAHARA, M.D., SHUNICHI NEGORO, M.D., TAKAHIKO SUGIURA, M.D., AKIRA YOKOYAMA, M.D., MASAHIRO FUKUOKA, M.D., KIYOSHI MORI, M.D., KOSHIRO WATANABE, M.D., TOMOHIDE TAMURA, M.D., SEIICHIRO YAMAMOTO, PH.D., AND NAGAHIRO SAIJO, M.D., FOR THE JAPAN CLINICAL ONCOLOGY GROUP*

N Engl J Med, Vol. 346, No. 2 · January 10, 2002 · www.nejm.org



No. at Risk							
Irinotecan plus cisplatin	77	67	45	21	15	11	7
Etoposide plus cisplatin	77	60	29	8	4	4	3

Figure 1. Overall Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.



No. at Risk				
Irinotecan plus cisplatin	77	47	9	9
Etoposide plus cisplatin	77	27	6	2

Figure 2. Progression-free Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.

Próba kliniczna SCLC (2)

IRINOTECAN PLUS CISPLATIN COMPARED WITH ETOPOSIDE PLUS CISPLATIN FOR EXTENSIVE SMALL-CELL LUNG CANCER

KAZUMASA NODA, M.D., YUTAKA NISHIWAKI, M.D., MASAAKI KAWAHARA, M.D., SHUNICHI NEGORO, M.D., TAKAHIKO SUGIURA, M.D., AKIRA YOKOYAMA, M.D., MASAHIRO FUKUOKA, M.D., KIYOSHI MORI, M.D., KOSHIRO WATANABE, M.D., TOMOHIDE TAMURA, M.D., SEIICHIRO YAMAMOTO, PH.D., AND NAGAHIRO SAIJO, M.D., FOR THE JAPAN CLINICAL ONCOLOGY GROUP*

N Engl J Med, Vol. 346, No. 2 · January 10, 2002 · www.nejm.org

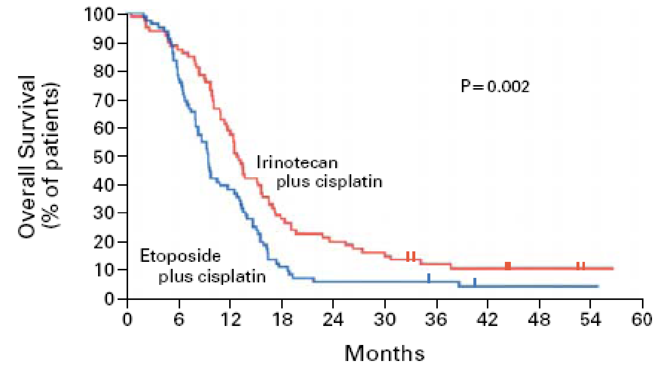
All comparisons of patients' characteristics, prognostic variables, response rates, and rates of toxic effects were performed with Fisher's exact test, except for age, for which the t-test was used. Survival was measured as the date of randomization to the date of death or the date of the most recent follow-up. Progression-free survival was measured as the date of randomization to the date of the first observation of disease progression or the date of death from any cause if there had been no progression. If there was no progression and if the patient had not died, data on progression-free survival were censored as of the date that the absence of progression was confirmed. If a patient died without information on progression, data on progression-free survival were censored as of the last date on which progression could be ruled out by review of follow-up forms. Survival curves were calculated by the Kaplan–Meier method[10] and compared with use of the log-rank test.

Próba kliniczna SCLC (3)

Overall Survival

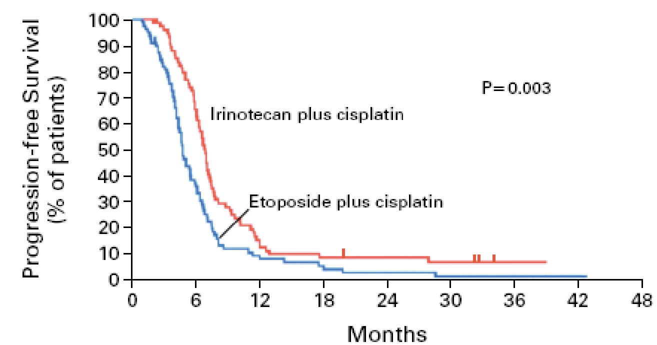
As of March 2001, when the final analysis was conducted, the median overall survival was 12.8 months (95 percent confidence interval, 11.7 to 15.2) in the irinotecan-plus-cisplatin group and 9.4 months (95 percent confidence interval, 8.1 to 10.8) in the etoposide-plus-cisplatin group; 70 patients in the irinotecan-

1). The rate of overall survival in the irinotecan-plus-cisplatin group was 58.4 percent (95 percent confidence interval, 47.4 to 69.4 percent) at one year and 19.5 percent (95 percent confidence interval, 10.6 to 28.3 percent) at two years; in the etoposide-plus-cisplatin group, the rates of overall survival at these time points were 37.7 percent (95 percent confidence interval, 26.8 to 48.5 percent) and 5.2 percent (95 percent confidence interval, 0.2 to 10.2 percent). The



No. at Risk							
Irinotecan plus cisplatin	77	67	45	21	15	11	7
Etoposide plus cisplatin	77	60	29	8	4	4	3

Figure 1. Overall Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.



No. at Risk				
Irinotecan plus cisplatin	77	47	9	9
Etoposide plus cisplatin	77	27	6	2

Figure 2. Progression-free Survival of Patients with Extensive Small-Cell Lung Cancer Who Were Assigned to Treatment with Irinotecan plus Cisplatin or Etoposide plus Cisplatin.
The tick marks indicate patients whose data were censored.

- Metoda użyta do obliczenia przedziałów ufności dla p-stwa przeżycia?

Szacowanie funkcji przeżycia dla danych ucinanych lewostronnie i cenzurowanych prawostronnie

- Czas włączenia do próby L_i; czas zdarzenia/cenzurowania T_i
- Obserwowane czasy zdarzeń $t_{(1)} < \ldots < t_{(m)}$
  - d_j – liczba zdarzeń dla $t_{(j)}$;
- „Uaktualniony" zbiór ryzyka dla $t_{(j)}$: obserwacje z $L_i < t_{(j)} \leq T_i$
  - Bez ucinania: obserwacje z $t_{(j)} \leq T_i$
- Estymator Kaplana-Meiera z „uaktualnionym" zbiorem ryzyka
  - Szacujemy $P(T^* \geq t)/P\{T^* \geq \min(L_i)\}$

Szacowanie funkcji przeżycia dla danych cenzurowanych lewostronnie

- „Odwrócenie” skali czasu: $T'=\tau-T^*$, gdzie τ „duże”
  - T' cenzurowane prawostronnie
- Używamy estymatora Kaplana-Meiera dla T'
  - Szacujemy $P(T' \geq t) = P(\tau-T^* \geq t) = P(T^* \leq \tau-t)$

Szacowanie funkcji przeżycia dla danych cenzurowanych lewo- lub prawostronnie (1)

- Procedura iteracyjna

- „Siatka" obserwowanych czasów $0=t_0 < t_1 <\dots < t_m$
  - d_j – liczba zdarzeń dla t_j;
  - r_j – liczba obserwacji prawostronnie cenzurowanych
  - l_j – liczba obserwacji lewostronnie cenzurowanych

Szacowanie funkcji przeżycia dla danych cenzurowanych lewo- lub prawostronnie (2)

- Krok 0: początkowy estymator $S_0(t_j)$
- Krok K+1:
  - szacujemy $p_{ij}=P(t_{j-1} \le T^* < t_j) | T^* < t_i)$ przez $\{S_K(t_{j-1})-S_K(t_j)\}/\{1-S_K(t_i)\}$, $j \le i$
  - szacujemy d_j przez $d_j+\sum^{m}_{i=j} (l_i\, p_{ij})$
  - estymator $S_{K+1}(t)$ to estymator Kaplana-Meiera zastosowany do oszacowań d_j i zaobserwowanych r_j
  - jeśli $S_{K+1}(t)$ jest „bliski" $S_K(t)$ dla każdego $t=t_j$, stop; w p.p., krok K+2

Szacowanie funkcji przeżycia dla danych cenzurowanych obustronnie (1)

- Dane w postaci $(L_i,R_i]$
- Procedura iteracyjna
- „Siatka” czasów $0=t_0 < t_1 < \ldots < t_m$, włącznie z L_i i R_i
- $a_{ij} = 1$ jeśli $(t_{j-1}, t_j] \subset (L_i,R_i]$, 0 w p.p.

Szacowanie funkcji przeżycia dla danych cenzurowanych obustronnie (2)

- Krok 0: początkowy estymator $S_0(t_j)$
- Krok K+1:
  - szacujemy p-stwo zdarzenia dla t_j, $p_j = \{S_K(t_{j-1}) - S_K(t_j)\}$
  - szacujemy liczbę zdarzeń dla t_j: $d_j = \sum_i a_{ij} p_j / \sum_k a_{ik} p_k$
  - szacujemy zbiór ryzyka dla t_j: $n_j = \sum^m_{k=j} d_k$
  - estymator $S_{K+1}(t)$ to estymator Kaplana-Meiera zastosowany do oszacowań d_j i n_j
  - jeśli $S_{K+1}(t)$ jest „bliski" $S_K(t)$ dla każdego $t=t_j$, stop; w p.p., krok K+2

Szacowanie funkcji przeżycia dla danych ucinanych prawostronnie

- Czas włączenia do próby L_i ; czas zdarzenia T_i
- Okres obserwacji τ
  - $L_i + T_i \leq \tau \rightarrow T_i \leq \tau - L_i$
- „Odwrócenie” skali czasu: $T' = \tau - T^*$
  - T' ucinane lewostronnie: $L \leq T'$
- Używamy estymatora dla T'
  - Szacujemy $P(T' \geq t \mid T' \geq 0) = P(T^* \leq \tau - t \mid T^* \leq \tau)$