Maastricht UMC+

Survival analysis:
van Edward Kaplan & Paul Meier tot David Cox

Bram Ramaekers
KEMTA Masterclass

“I’ve survived a lot of things, and I’ll probably survive this.”

J.D. Salinger

Maastricht UMC+

Inhoud

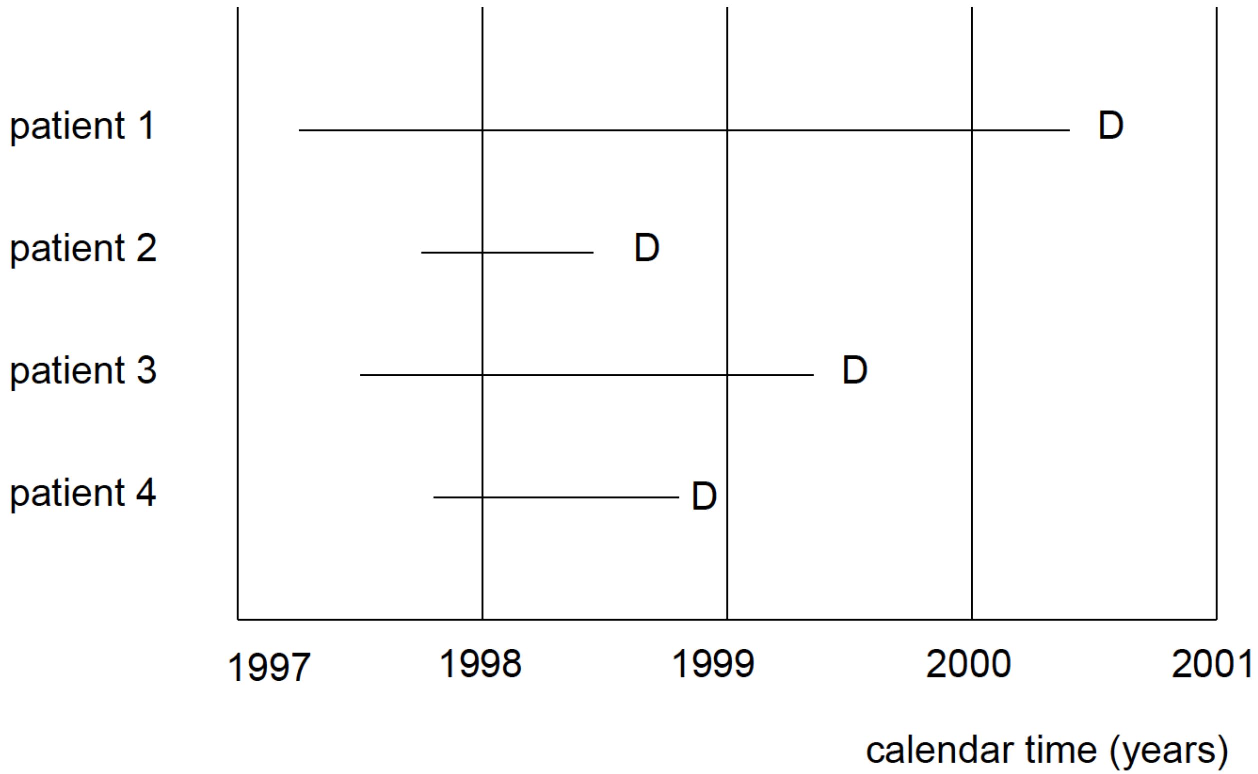
- Data
- Kaplan-Meier curve
- Log-rank test
- Hazard rate en hazard ratio
- Cox regressie

Beginpunt Tijdsduur **Eindpunt**

Data

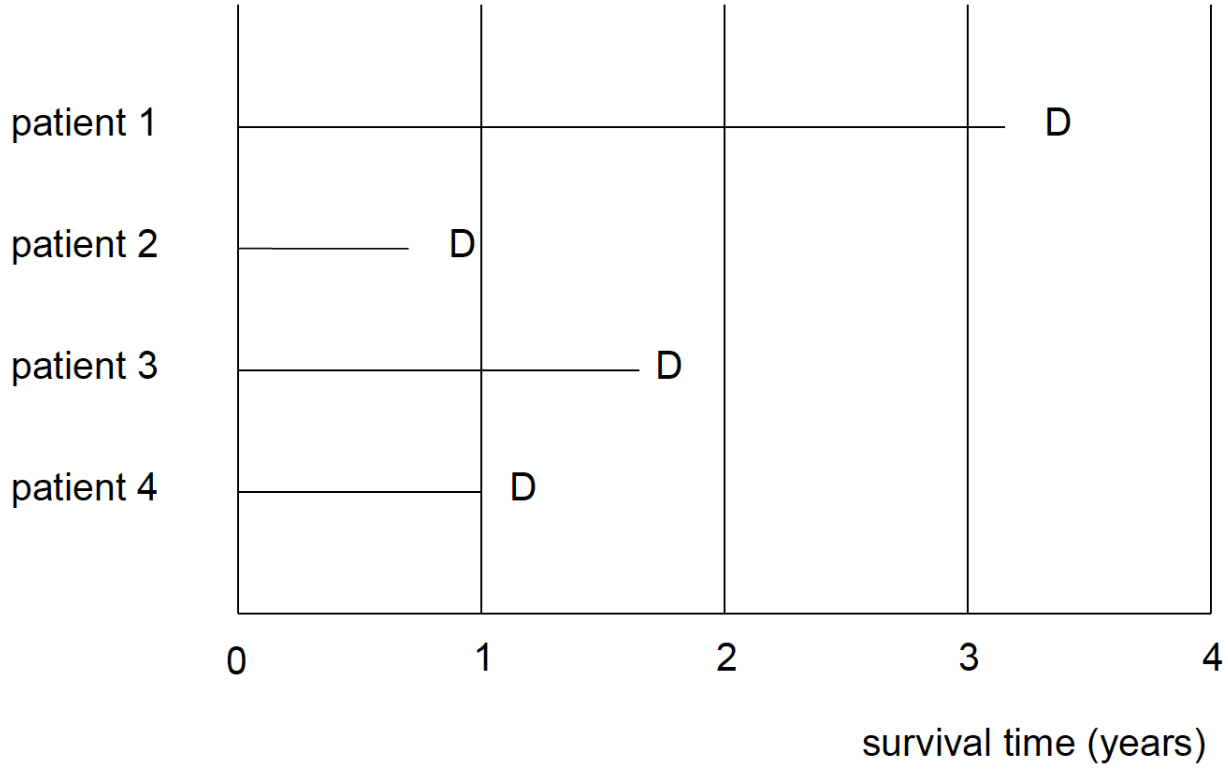
Tijd tot event

Data



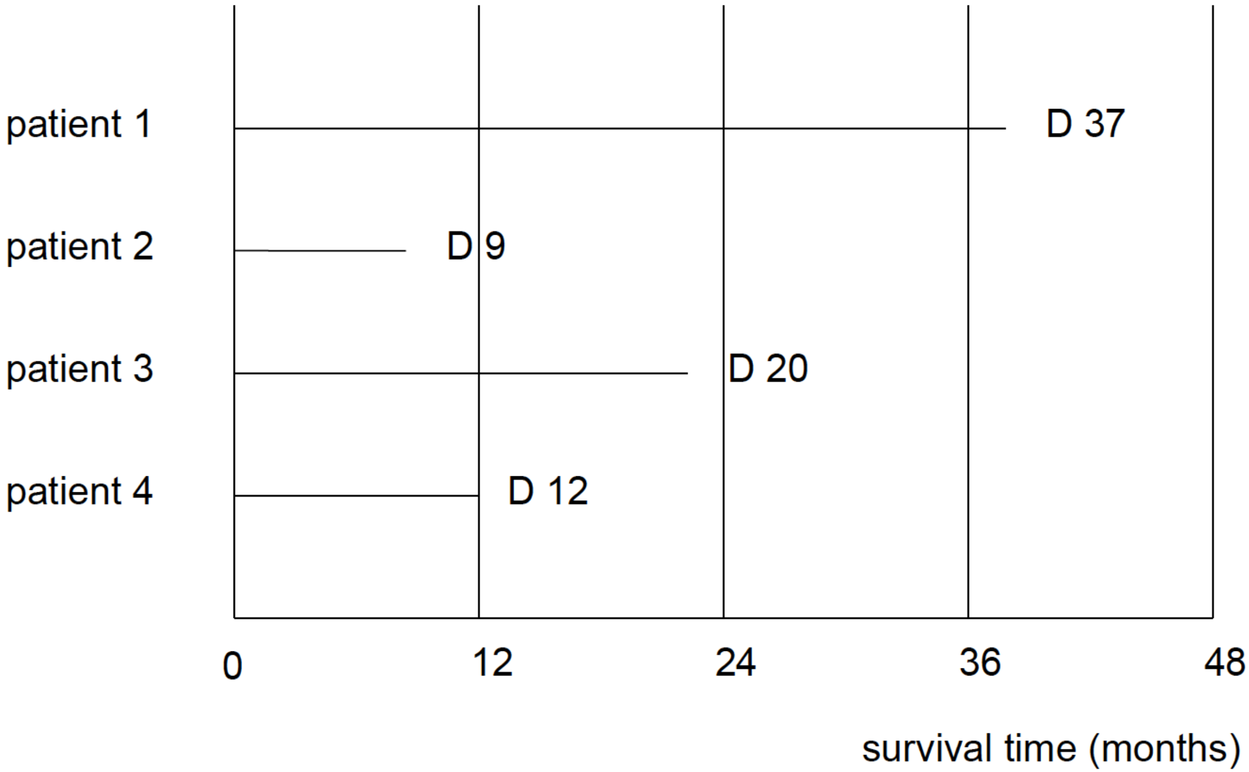
Slides based on material of Ton Ambergen

Data



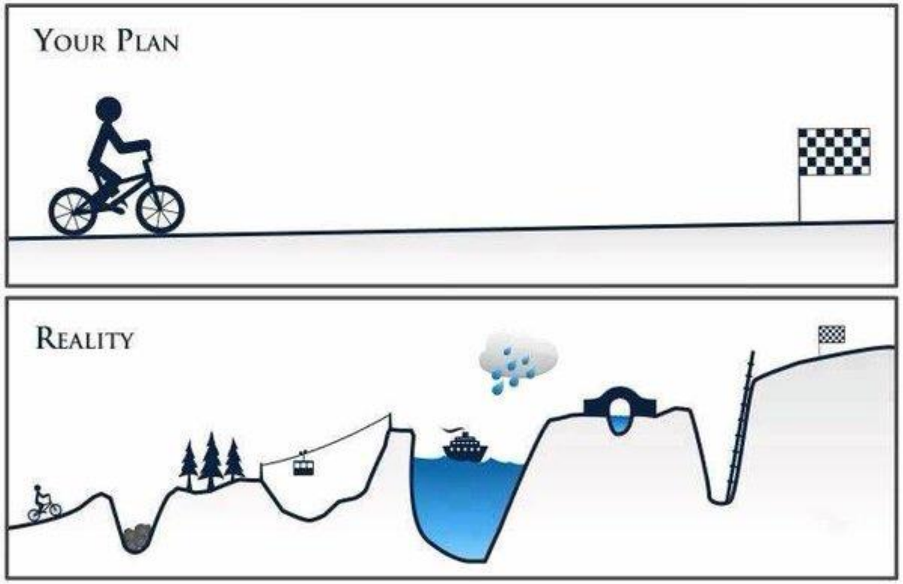
Slides based on material of Ton Ambergen

Data

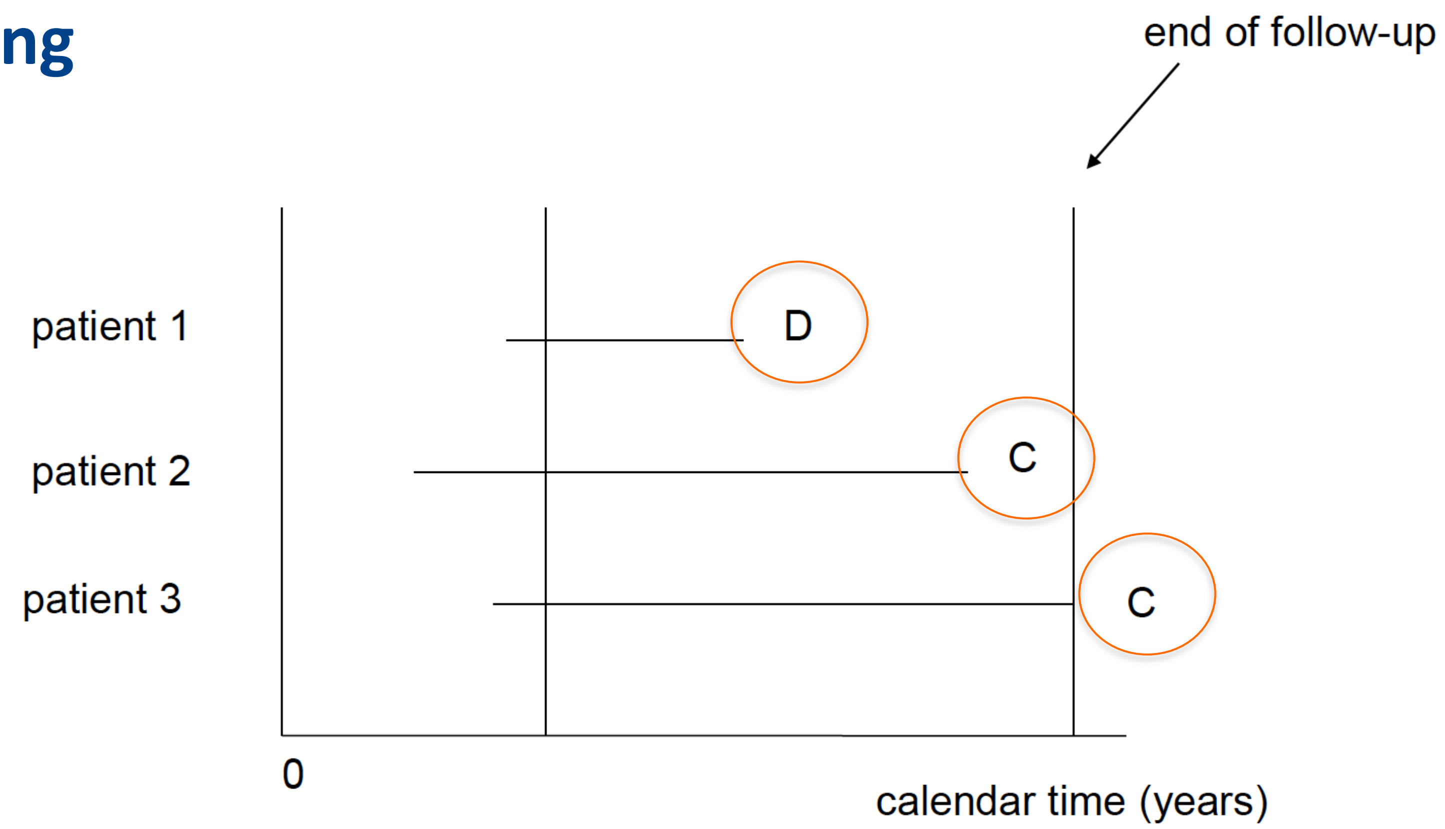


Slides based on material of Ton Ambergen

Verwachting versus werkelijkheid



Censoring



D = death
C = censored

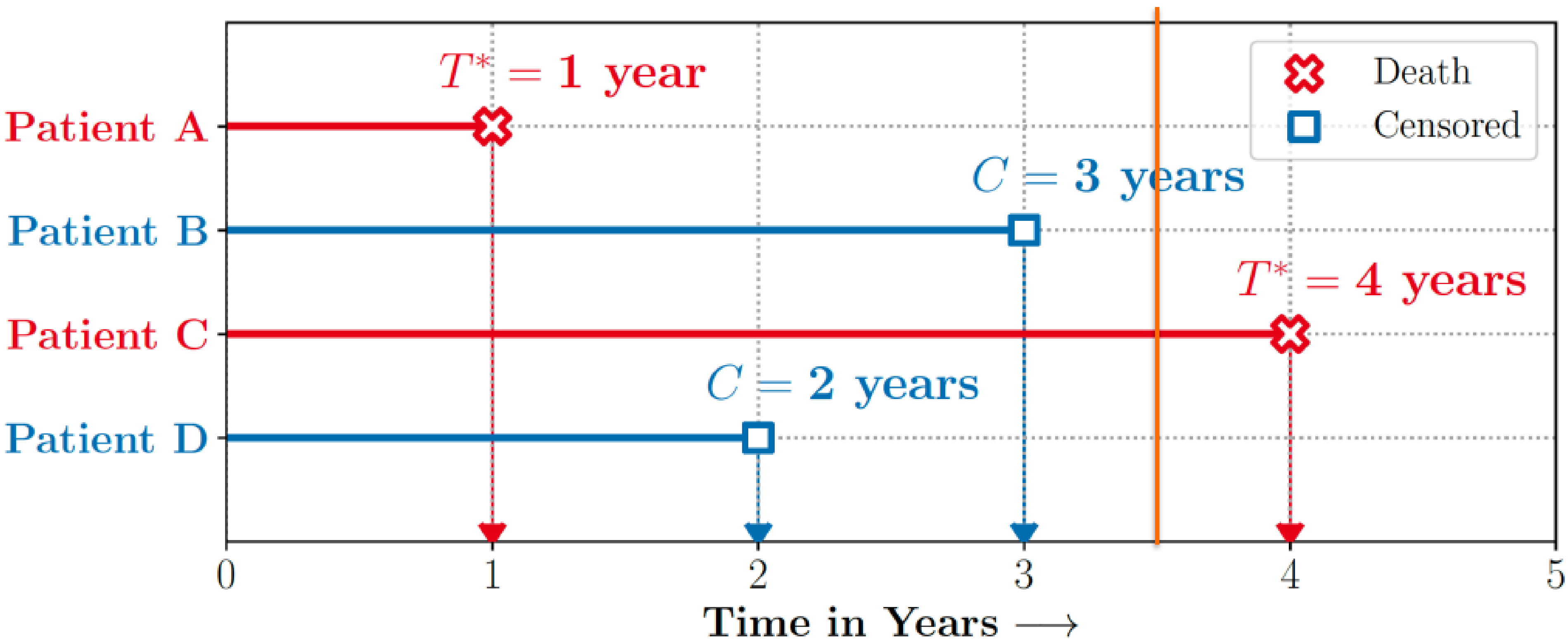
Slides based on material of Ton Ambergen

Hoe om te gaan met censored cases?

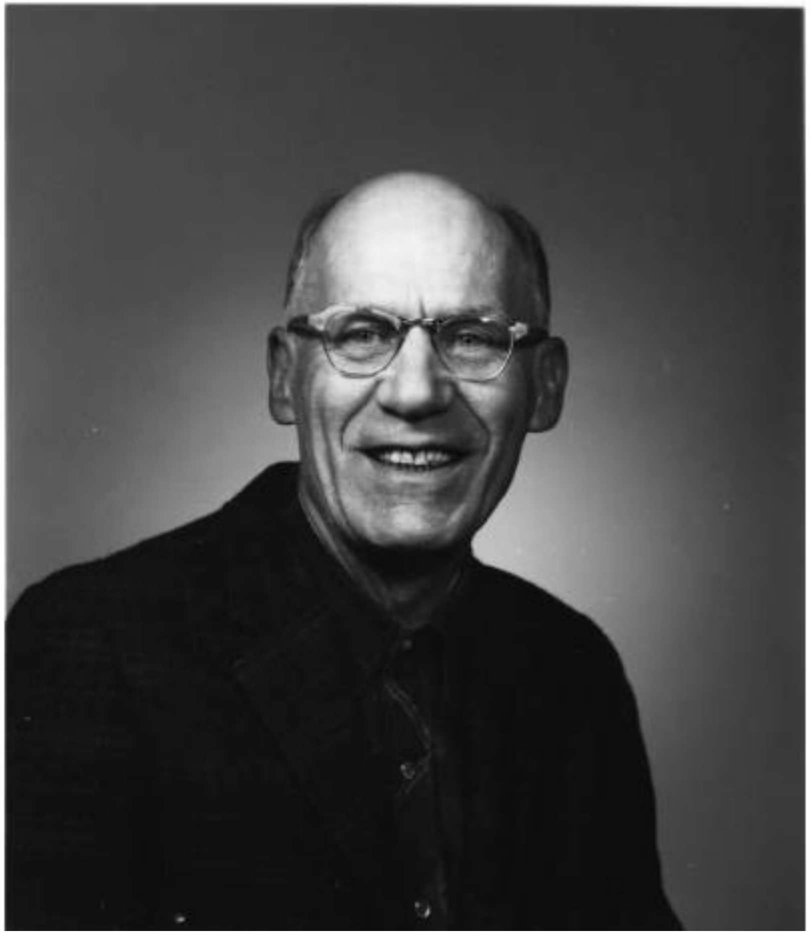
Kans op overlijden na 3.5 jaar (oranje lijn)?

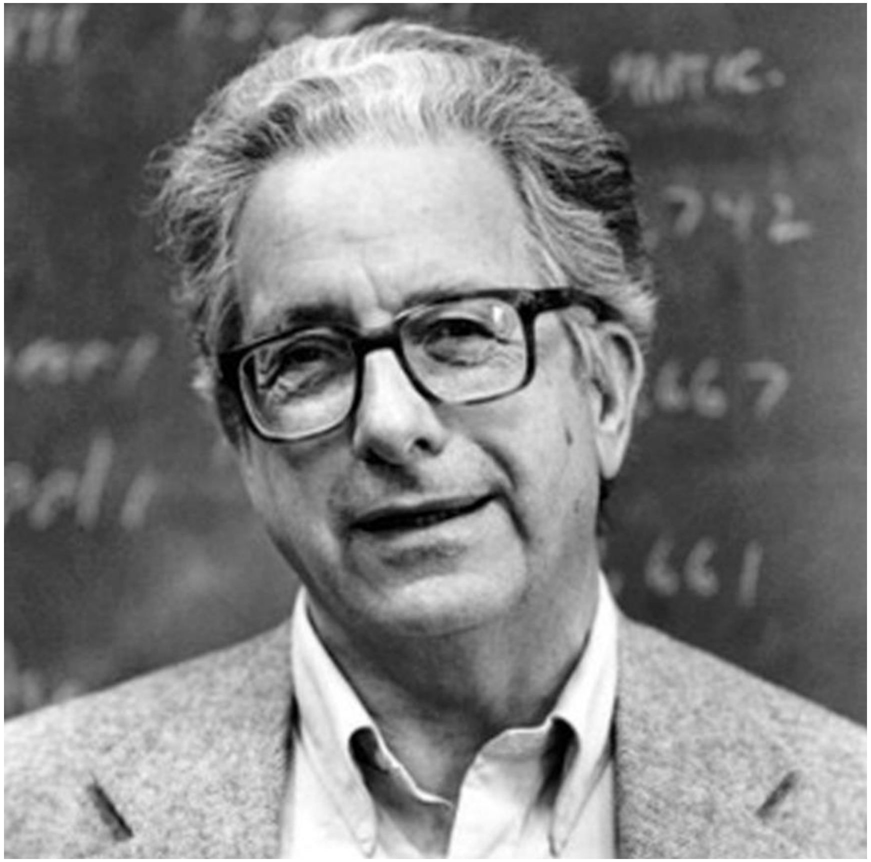
- 1 dood (A), 2 censored (B en D), 1 overleefd (C)

    I) 3/4 II) 1/4 III) 1/2



Kaplan-Meier analyse





Kaplan-Meier analyse

Idee:

Om *t* tijdseenheden te overleven, moet je de eerste *t-1* eenheden hebben overleefd en dan ook de *t*[e]

S(t)=S(t-1) · P(surviving t[th] time unit)

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

Kaplan-Meier analyse

Wagenziekte

Max 2 uur

Time to event:

- 1e keer overgeven

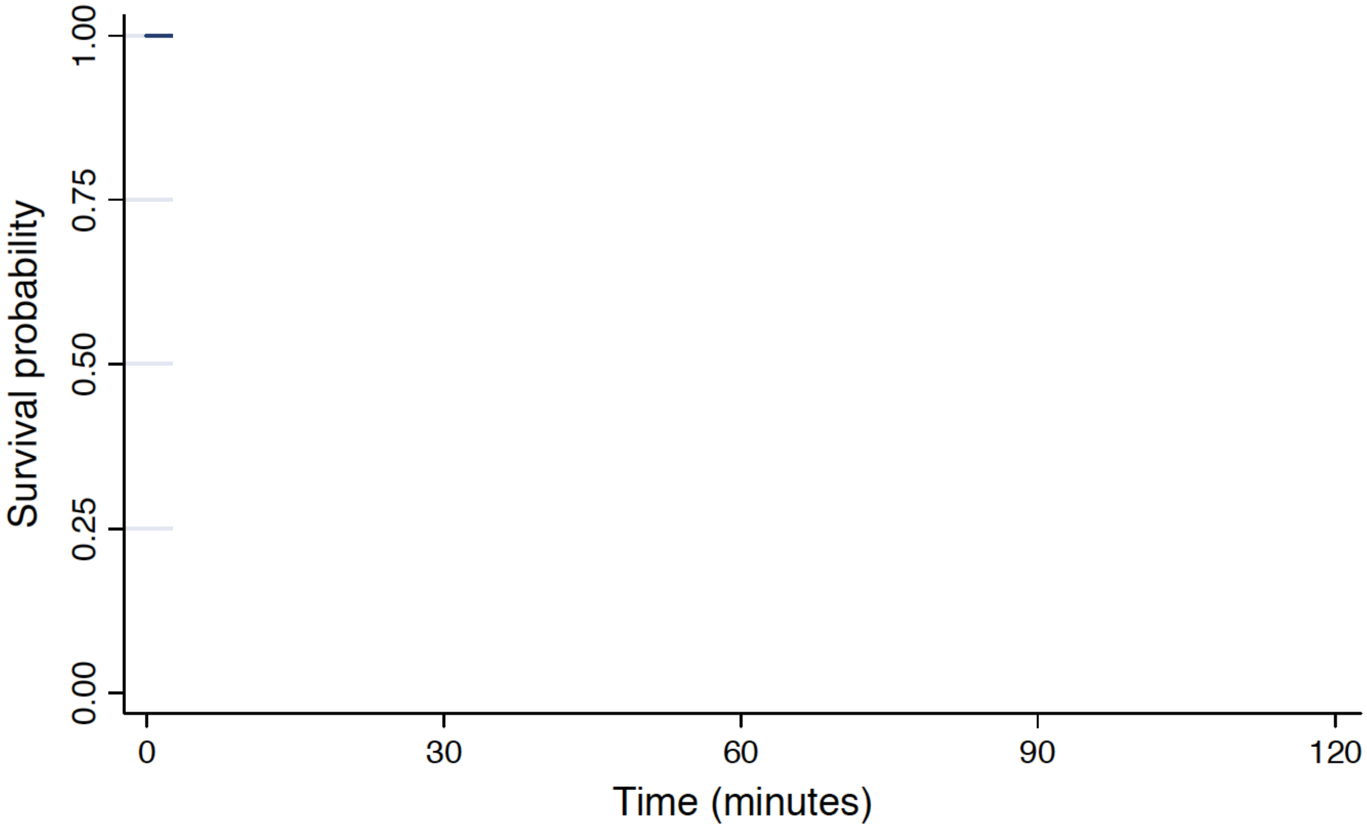
Subject	Time (minutes)	Event (1=yes)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	1
7	92	1
8	120	0
...	...	...
21	120	0

Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	$1\cdot 21/21=1$

Kaplan-Meier curve

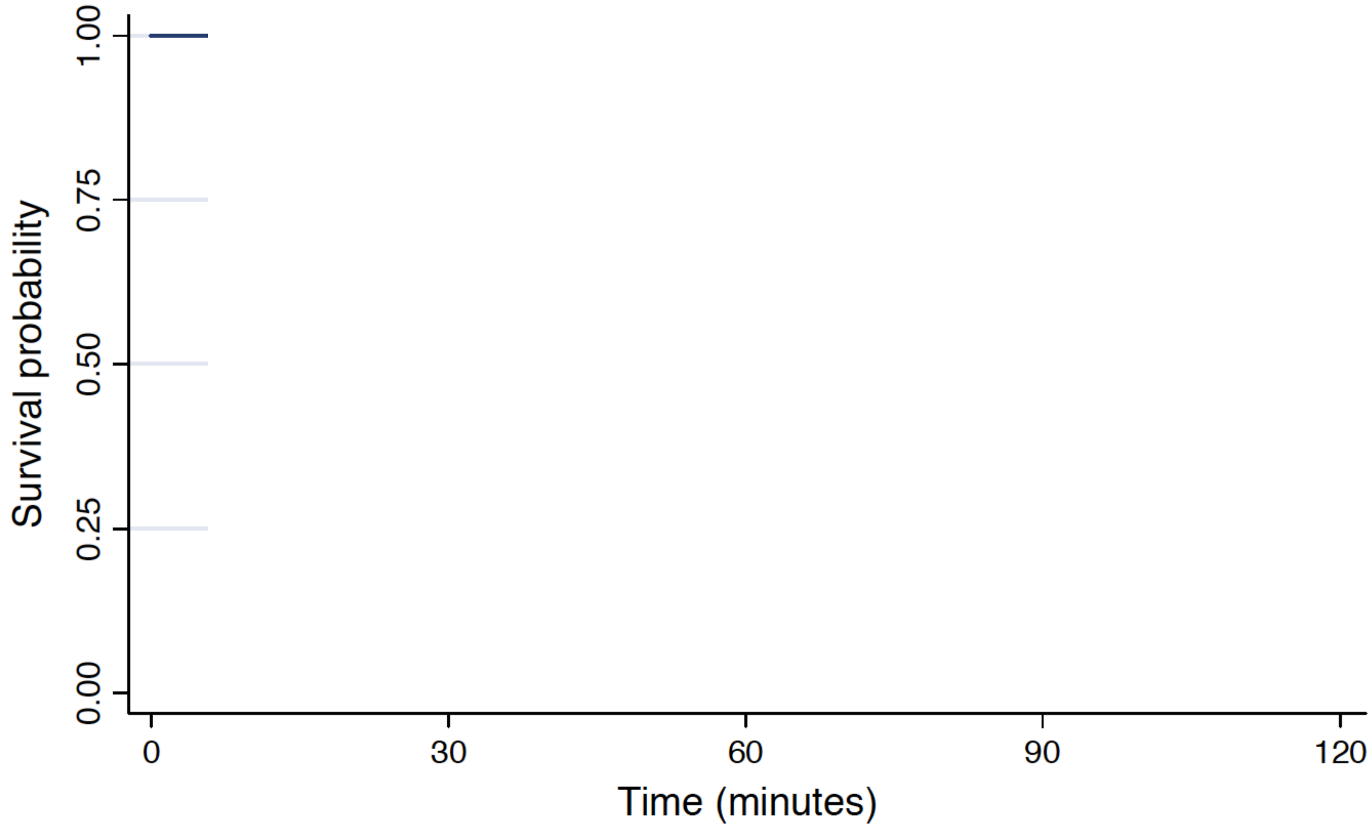


Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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

Kaplan-Meier curve

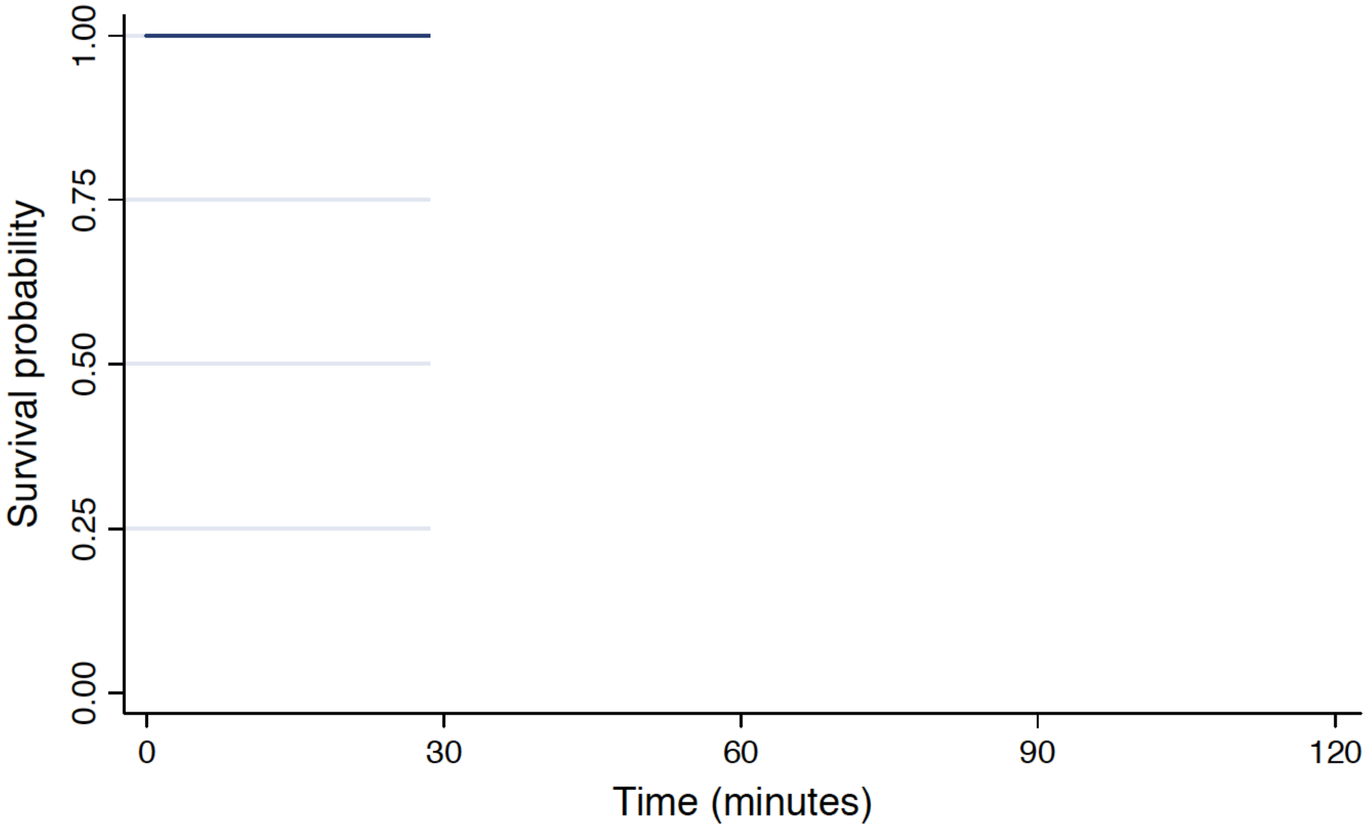


Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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

Kaplan-Meier curve

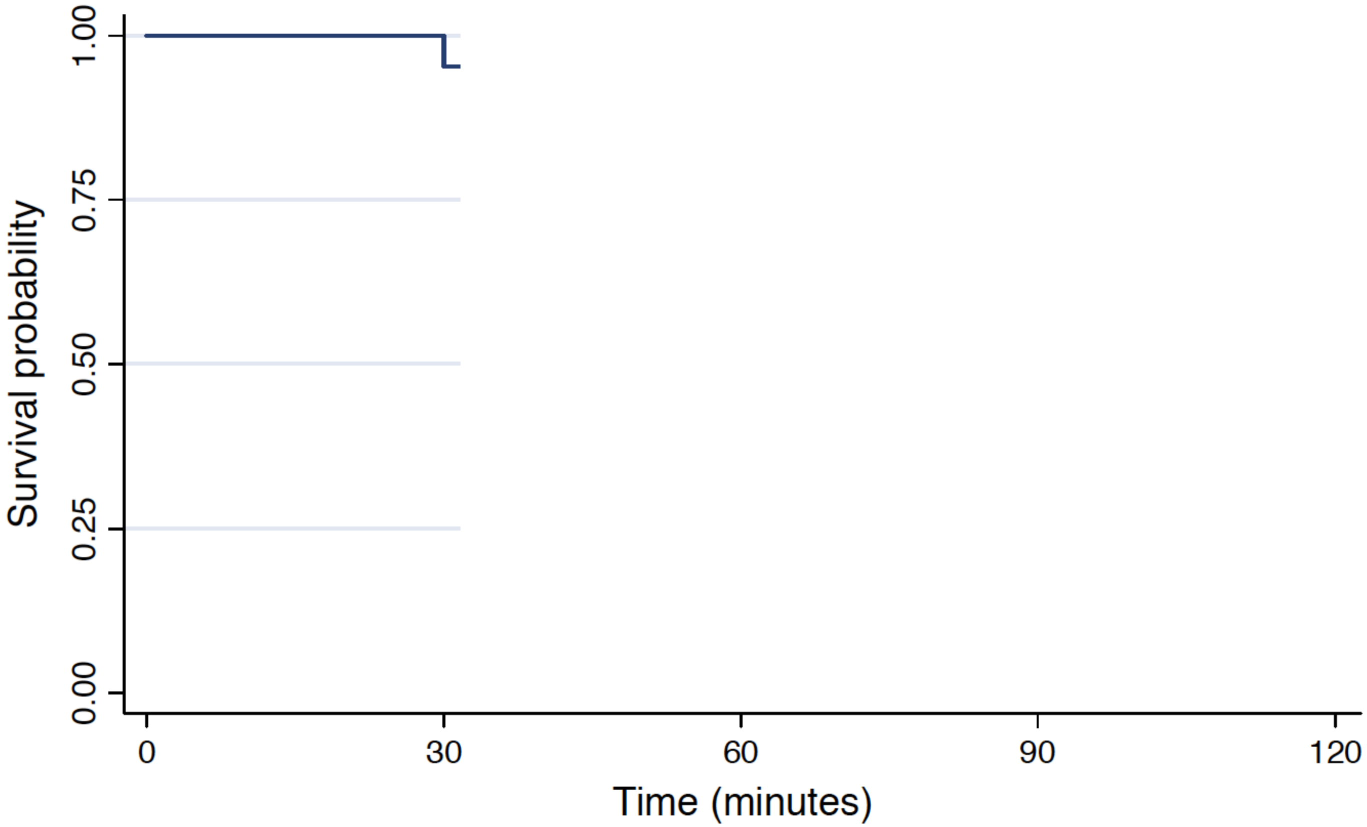


Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 0-t) $S(t)=S(t-1)\cdot P_t$
1	21	0	21/21	$1\cdot 21/21=1$
2	21	0	21/21	$1\cdot 21/21=1$
…	21	0	21/21	$1\cdot 21/21=1$
29	21	0	21/21	$1\cdot 21/21=1$
30	21	1	20/21	$1\cdot 20/21=0.952$

Kaplan-Meier curve

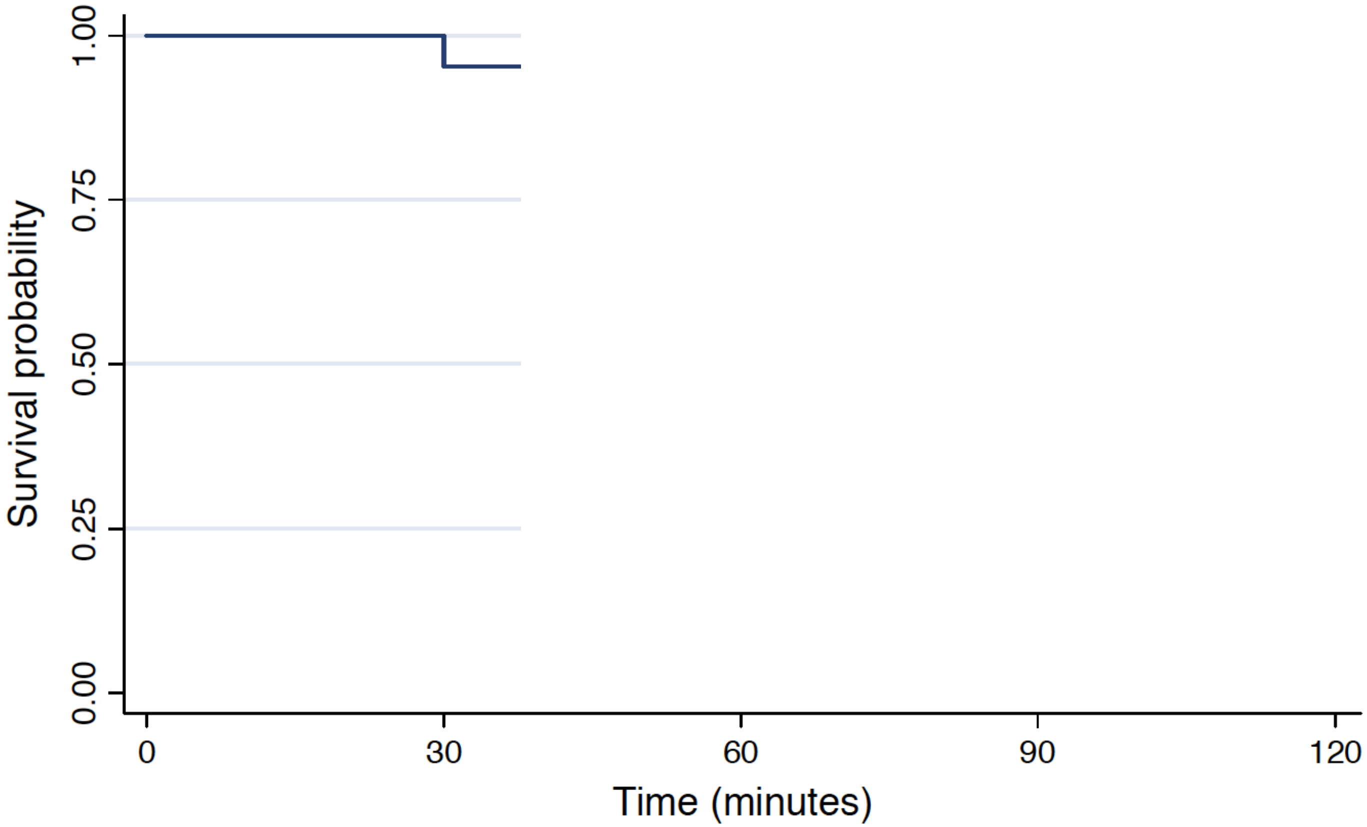


Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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

Kaplan-Meier curve

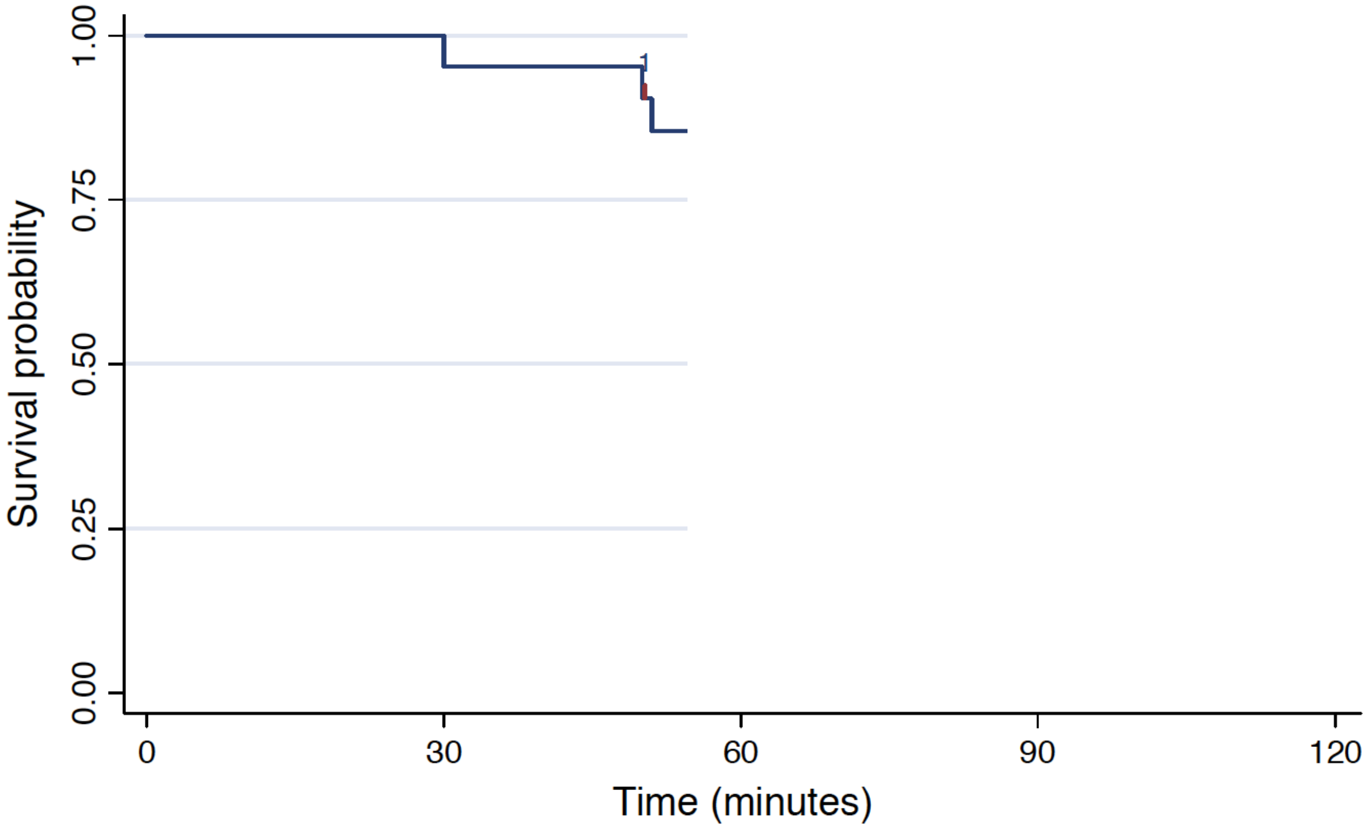


Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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

Kaplan-Meier curve



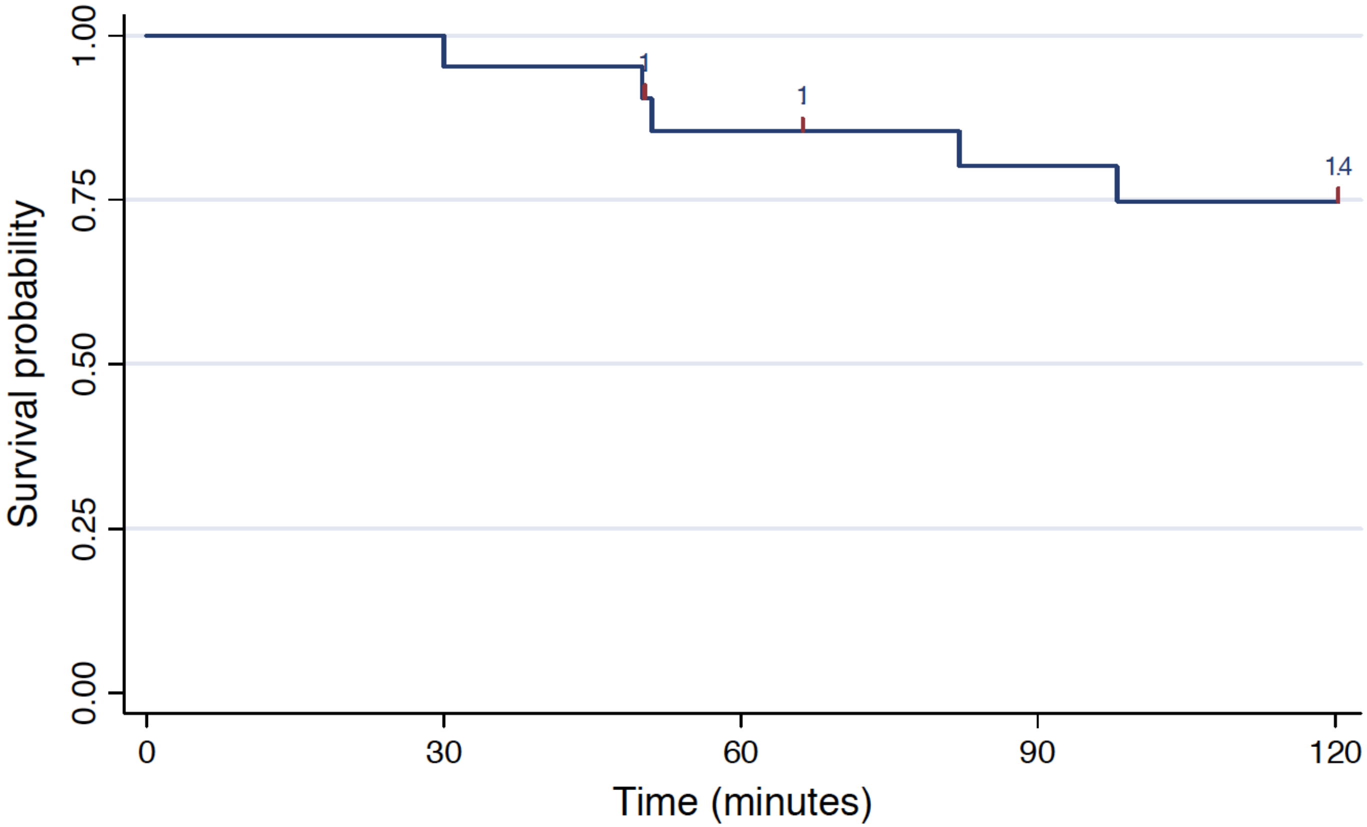
Maastricht UMC+

Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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
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

Kaplan-Meier curve



Kaplan-Meier analyse

Subject	Time (min)	Event (1=yes)
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

Time t	Risk-set at t R_t	Events at t D_t	Survival (0 events) of t $P_t=(R_t-D_t)/R_t$	Survival function (0 events in 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
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

Kaplan-Meier analyse

Subject	Time (minutes)	Event (1=yes)
1	30	1
2	50	1
3	50	0
4	51	1
5	66	0
6	82	[illegible]
7	92	[illegible]
8	120	[illegible]
...	...	...
21	120	0

Time t	Risk set (alive) at t R_t	Events at t D_t	Probability surviving t $P_t=(R_t - D_t)/R_t$	Survival probability (no event in 0-t) $S(t)=S(t\text{-}1)\cdot P_t$
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

Aanname: tijd tot event moet onafhankelijk zijn van tijd tot censoring
Dus censoring is onafhankelijk van prognosis (indien dood het event is)

Voorbeeld

Dataset borstkanker patienten (n=457):

Twee groepen:
- Interventie (n=229)
- Controle (n=228)

Tijdsduur:
Mean 2.98 (1.765), median 2.67 (0-7)

Overlijden:
n=196 (42.9%)

Survival tabel SPSS output

Means and Medians for Survival Time

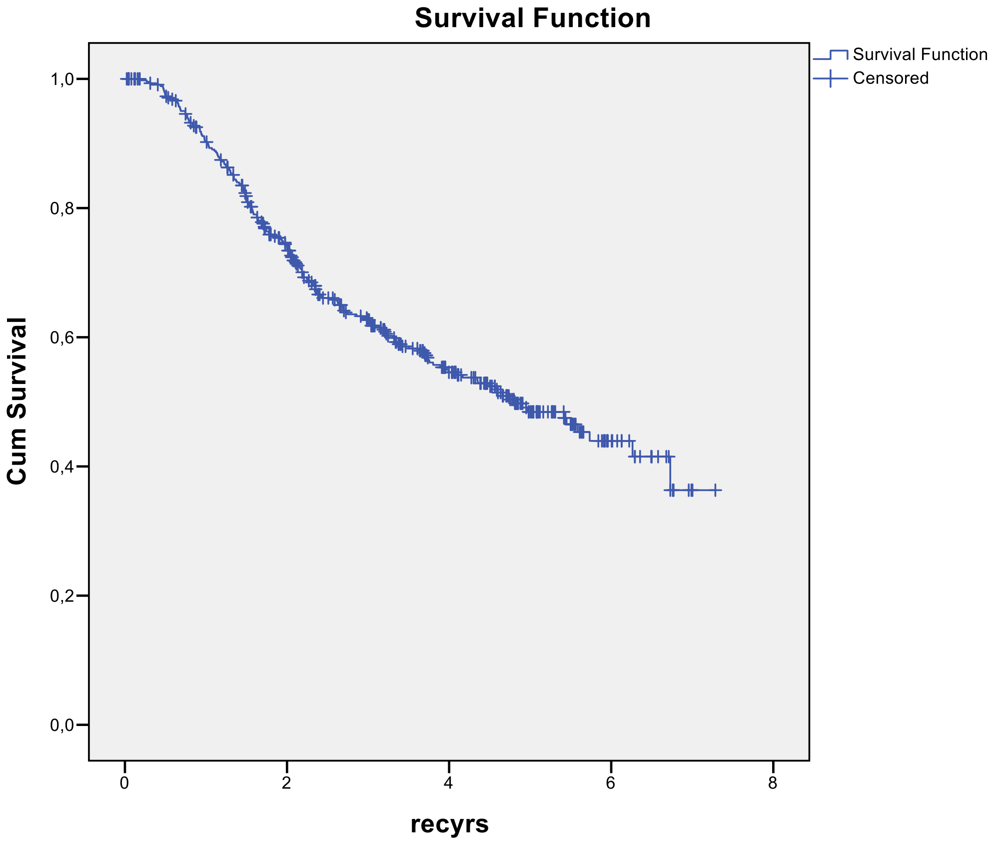
Mean [a]				Median			
		95% Confidence Interval				95% Confidence Interval	
Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
4,538	,141	4,261	4,815	4,803	,453	3,915	5,690

a. Estimation is limited to the largest survival time if it is censored.

Survival Table

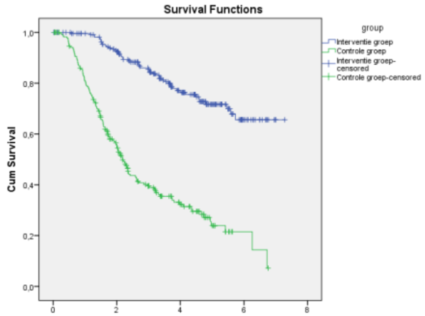
	Time	Status	Cumulative Proportion Surviving at the Time		N of Cumulative Events	N of Remaining Cases
			Estimate	Std. Error		
1	,022	0	.	.	0	456
2	,041	0	.	.	0	455
3	,044	0	.	.	0	454
4	,047	0	.	.	0	453
5	,049	0	.	.	0	452
6	,079	0	.	.	0	451
7	,115	0	.	.	0	450
8	,126	0	.	.	0	449
9	,156	0	.	.	0	448
10	,178	0	.	.	0	447
11	,184	0	.	.	0	446
12	,197	1	,998	,002	1	445
13	,268	1	,996	,003	2	444
14	,310	1	,993	,004	3	443
15	,312	0	.	.	3	442
16	,329	1	,991	,004	4	441
17	,405	0	.	.	4	440
18	,438	1	,989	,005	5	439
19	,468	1	,987	,005	6	438
20	,474	1	,984	,006	7	437
21	,479	1	,982	,006	8	436
22	,485	1	,980	,007	9	435
23	,493	1	,978	,007	10	434
24	,496	1	,975	,007	11	433

Kaplan-Meier curve



Kaplan-Meier curve

Statistisch significant verschillend?



Log-rank test (a.k.a. Mantel–Cox test)

- Vaak gebruikt (i.c.m. KM curves)
- Non-parametrisch
  - Gebruik voor rechts scheef + gecensureerde data
  - Zelfde aanname als KM: censureren niet gerelateerd aan prognose
- Test of er een verschil is in kans op event

Log-rank test

1. Per groep aantal geobserveerde en verwachte events

	Observed (O)	Expected (E)
Controle	14	22.48
Interventie	28	19.52

2. Per groep: $(O - E)^2 / E$ → dan optellen:
   $(14 - 22.48)^2 / 22.48 + (28 - 19.52)^2 / 19.52 = 6.88$
3. Aantal vrijheidsgraden = aantal groepen – 1
4. Opzoeken in Chi^2 verdeling tabel → $p < 0.01$

(Voorbeeld uit BMJ 2004;328:1073 https://doi.org/10.1136/bmj.328.7447.1073)

Log-rank test

Case Processing Summary

group	Total N	N of Events	Censored	
			N	Percent
Interventie groep	229	51	178	77,7%
Controle groep	228	145	83	36,4%
Overall	457	196	261	57,1%

Means and Medians for Survival Time

group	Mean [a]				Median			
	Estimate	Std. Error	95% Confidence Interval		Estimate	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound			Lower Bound	Upper Bound
Interventie groep	5,934	,162	5,618	6,251	.	.	.	.
Controle groep	3,064	,166	2,738	3,390	2,184	,116	1,956	2,411
Overall	4,538	,141	4,261	4,815	4,803	,453	3,915	5,690

a. Estimation is limited to the largest survival time if it is censored.

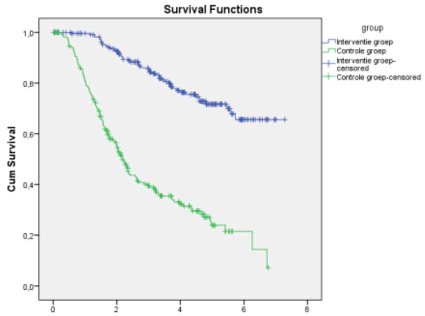
Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	113,591	1	,000

Test of equality of survival distributions for the different levels of group.

Kaplan-Meier curve

Statistisch significant verschillend?
Ja, p = 0.000



Hazard rate and hazard ratio?

- Iemand een idee?

Hazard function

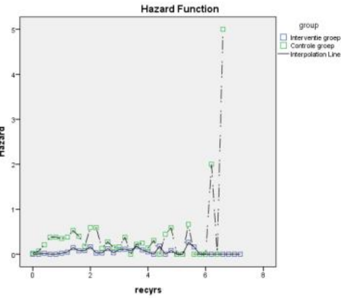
- Veel gebruikt in survival analysis

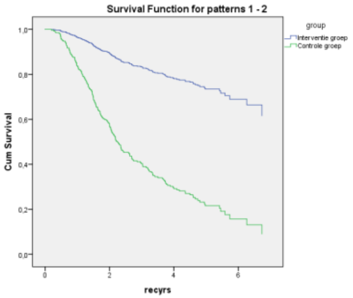
$$h(t) = \lim_{\delta \to 0} \frac{P(t \leq T < t + \delta | T > t)}{\delta}$$

Hazard rate

- Rate (per unit of time) of experiencing the event at time t, given that you are event free until that time
  - Intensiteit van het optreden van events, indien ‘at risk’
- Informatief voor het begrijpen van (natuurlijk) beloop van ziekte
- Veel gebruikt in survival analysis

Hazard rate

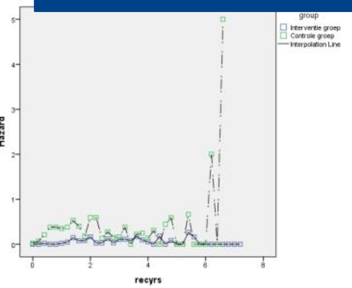


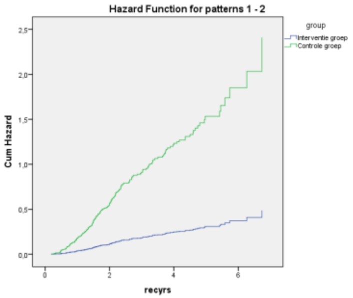


Maastricht UMC+

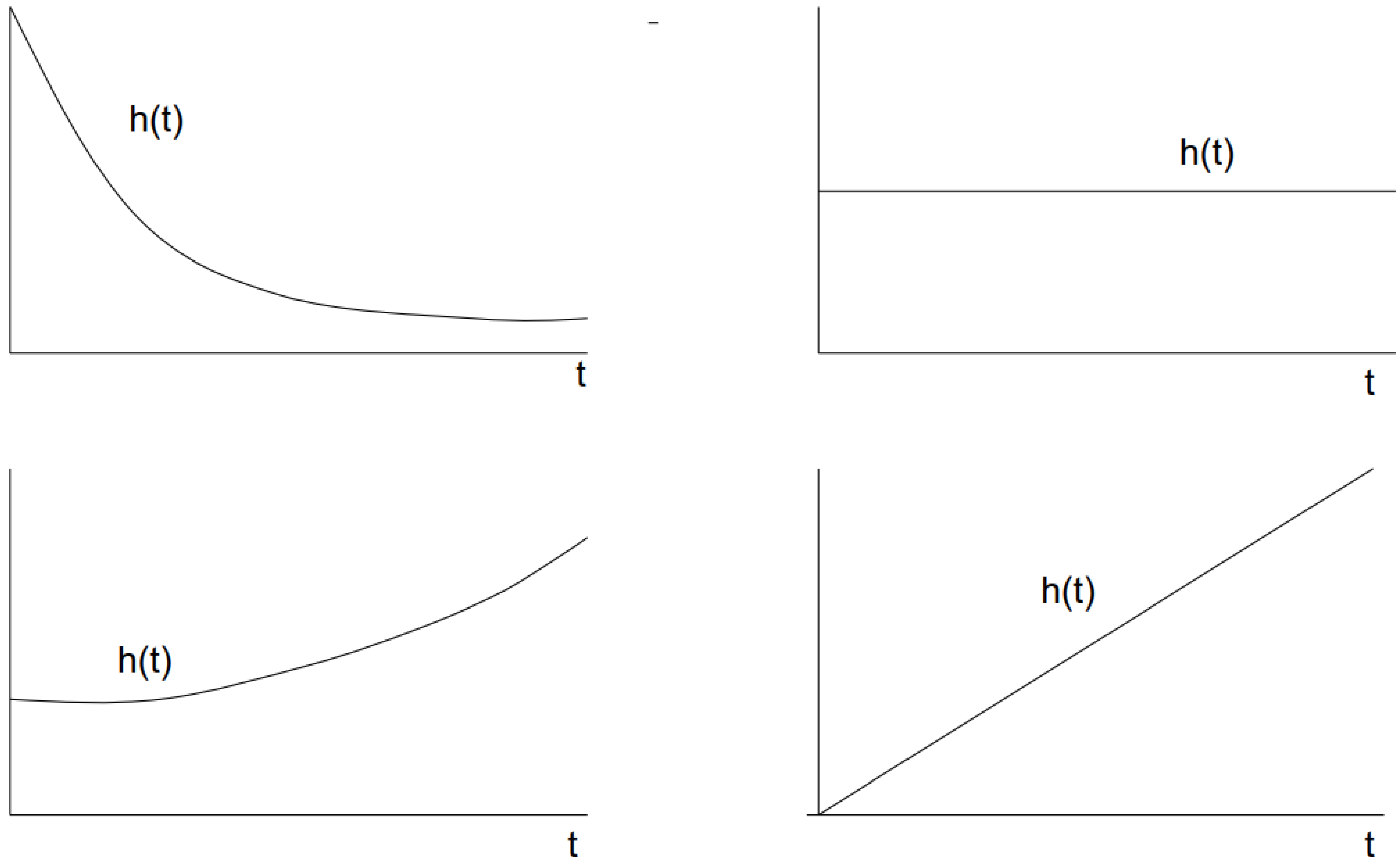
Hazard rate

Smoothed hazard is vaker eenvoudiger te interpreteren





Hazard rate - verschillende vormen



Hazard ratio

- Ratio van de hazard rates

  = hazard interventie groep / hazard controle groep
- Interpretatie van HR = 1?
- Verschil met RR en OR?
  - HR cumulatief over gehele studie periode
  - Ratio of hazards vs ratio of odds vs ratio of probabilities

Interpretatie van hazard ratio

- Interventie vs controle groep:
    - HR = 0.201 (95% CI: 0.146 – 0.279)
    - Interpretatie?

A) Op ieder tijdstip resulteert de interventie in risico reductie 80% t.o.v. controle groep (i.e. 80% minder kans op een event)

B) De interventie groep heeft 80% minder events op het einde van de follow-up

C) Absolute maat voor de tijdsduur zonder event

D) Rate of events (hazard rate) in de interventie groep is ~1 / 5 van de rate of events in de controle groep

E) De HR is statistisch significant

Interpretatie van hazard ratio

- Interventie vs controle groep:
  - HR = 0.201 (95% CI: 0.146 – 0.279)
  - Interpretatie?

A) **Op ieder tijdstip resulteert de interventie in risico reductie 80% t.o.v. controle groep (i.e. 80% minder kans op een event)**

B)

In het kort: op ieder tijdstip is de probability dat een individu een event heeft is 5 (=1/0.201) keer zo groot in de controle groep (t.o.v. de interventie groep)

C)

D) **Rate of events (hazard rate) in de interventie groep is ~1 / 5 van de rate of events in de controle groep**

E) **De HR is statistisch significant**

Cox regressie

- Relatief effect geschat (hazard rate $h_0(t)$ niet direct geschat)
    - Semi-parametrisch

$$h_i(t) = h_0(t) \exp(\phi^T \boldsymbol{v_i})$$

Cox regressie

- Relatief effect geschat (hazard rate $h_0(t)$ niet direct geschat)
    - Semi-parametrisch
- Pre: geen assumpties hazard rate
- Con: moeilijk om absolute voorspellingen / extrapolaties te doen
    - → parametrisch survival model
- Proportionaliteit van hazard rates (onafhankelijk van tijd)

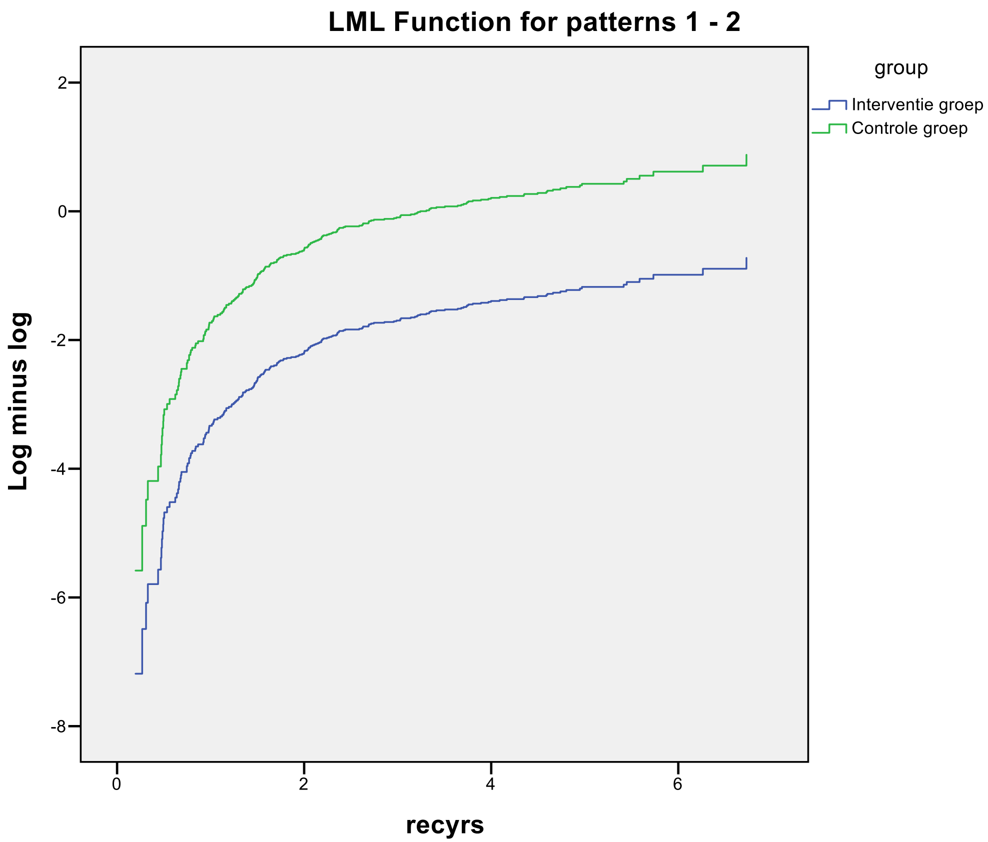
Cox regressie

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)	95.0% CI for Exp(B)	
							Lower	Upper
group	-1,602	,165	93,808	1	,000	,201	,146	,279

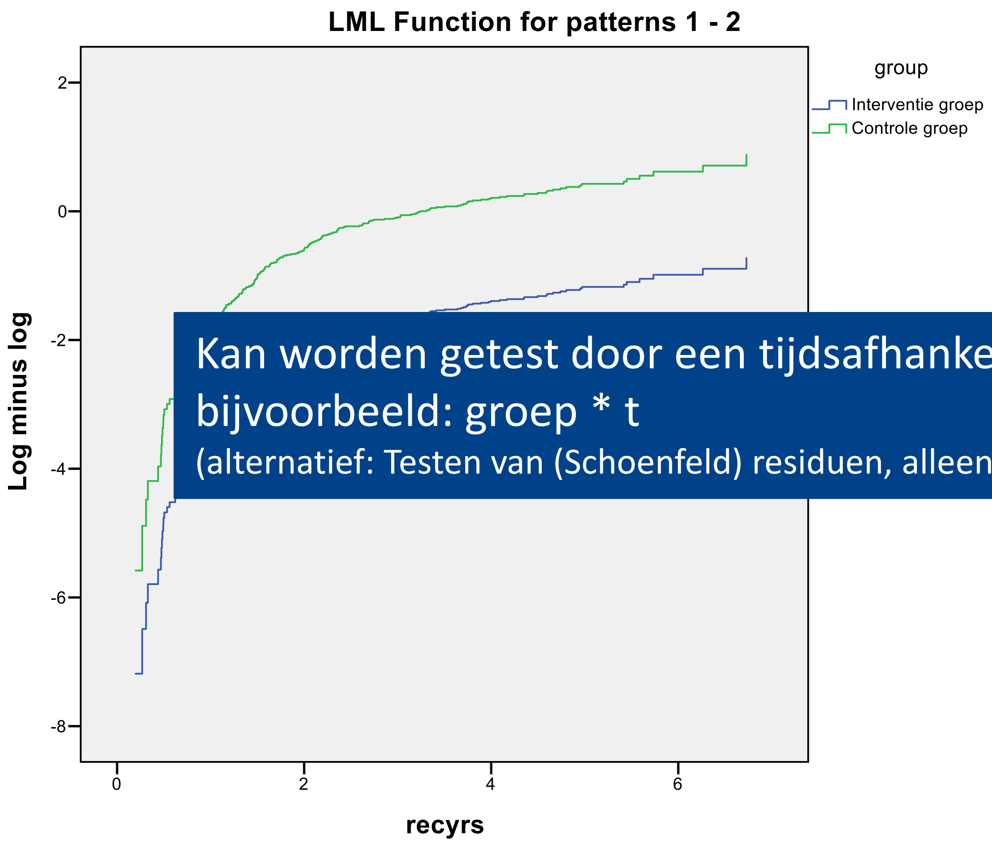
- Vergelijking groep (meerdere covariaten mogelijk)
- Interpretatie van de output?
- Hazard ratio = Exp(B) = 0.201

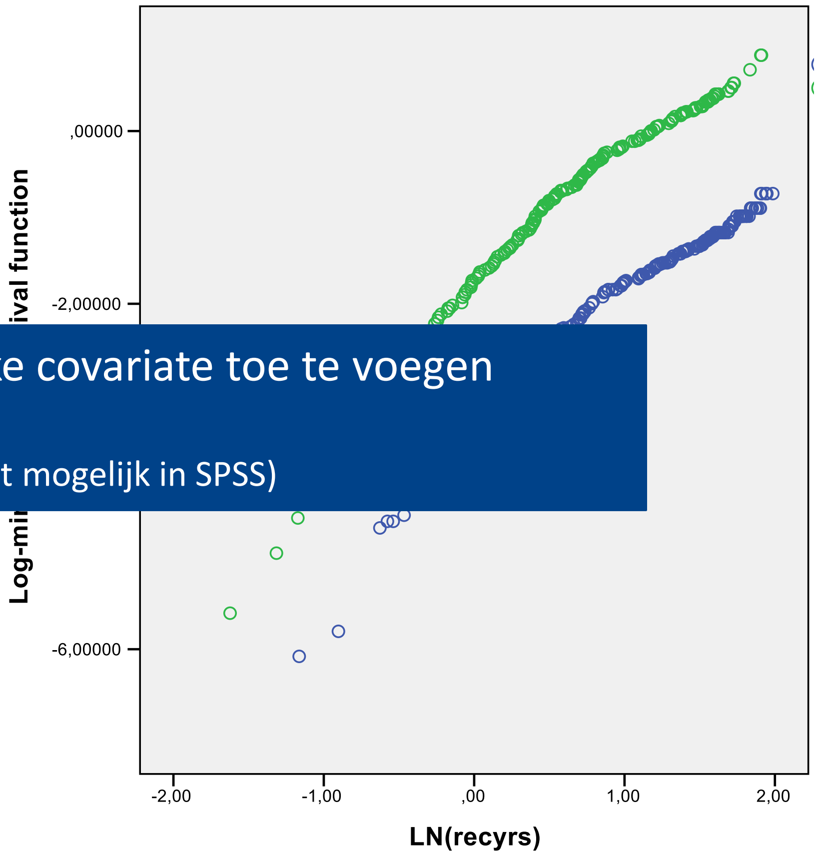
Proportionaliteit van hazards aanname



Note: -LN(survival) = cumulative hazard function

Proportionaliteit van hazards aanname





Kan worden getest door een tijdsafhankelijke covariate toe te voegen
bijvoorbeeld: groep * t
(alternatief: Testen van (Schoenfeld) residuen, alleen niet mogelijk in SPSS)

Note: -LN(survival) = cumulative hazard function

Strategieën bij niet-proportionele hazards

- Is het van belang (klinisch relevant)?
- Is het ‘echt’ (e.g. niet beïnvloed door uitschieters)?
- Indien beiden ja:
  - Deel de tijd as op in stukken
    - PH aanname is misschien gerechtvaardigd voor delen van de tijd periode
  - Gebruik tijds afhankelijke covariaten
    - Let op met data-driven keuzes
- Alternatief: interpreteer als ‘gemiddelde HR’ over tijd

Bedankt voor de aandacht

- Hands on practice: syntax + data gebruikt in presentatie kan worden opgevraagd
- Indien advies noodzakelijk kan contact op worden genomen met KEMTA

Vragen / opmerkingen