

ANALISI DELLA SOPRAVVIVENZA

Prof. Giuseppe Verlato

Sezione di Epidemiologia e Statistica Medica
Università degli Studi di Verona

OBIETTIVI DELL'ANALISI DELLA SOPRAVVIVENZA

- 1) **STIMARE** la funzione di sopravvivenza (ad esempio, probabilità cumulativa di sopravvivenza a 3 o 5 anni)
- 2) **CONFRONTARE** le esperienze di vita di gruppi di pazienti sottoposti a trattamenti diversi
- 3) **VALUTARE** la capacità prognostica di diverse variabili considerate separatamente e/o congiuntamente

Per effettuare un'analisi della sopravvivenza sono indispensabili due variabili:

- 1) La variabile **sensor (troncamento)** indica se il paziente non ha avuto l'evento terminale ($\text{sensor}=0$) o se ha avuto l'evento terminale ($\text{sensor}=1$) durante il periodo di osservazione.
- 2) La variabile **survival (sopravvivenza)** indica la durata del periodo di osservazione: **T1-T0**. Viene in genere espresso in anni o in mesi.

Dati necessari per l'analisi della sopravvivenza

Popolazione candidata = una **coorte (popolazione fissa)** di 6 pazienti diabetici, seguiti dall' 1-1-1997 e il 31-12-1999.



● = evento (vivo) → uscito vivo alla fine dello studio (withdrawn alive)
 X = evento (morto) → perso di vista prima della fine (lost to follow-up)

Paziente	Censor	Survival
1	0	3
2	0	3
3	0	3
4	0	2
5	1	1
6	1	3

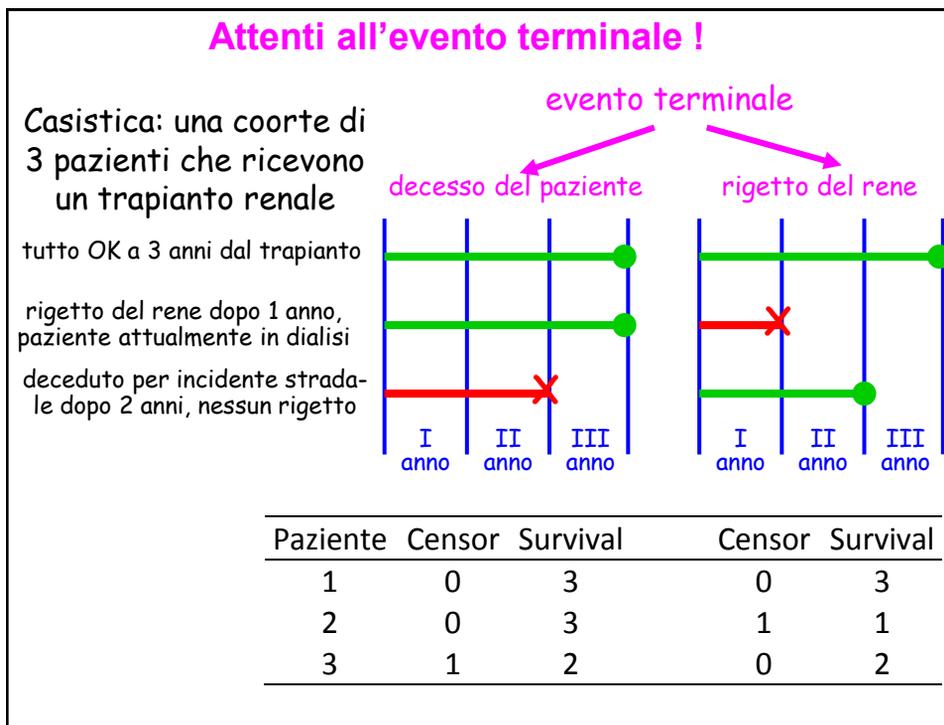
IL MOMENTO INIZIALE (TEMPO 0) può essere lo stesso per tutti i pazienti, ovvero un momento del calendario.

Tuttavia, nella maggior parte dei casi il tempo zero varia da paziente a paziente perché coincide con il momento della diagnosi o con l'intervento chirurgico o con l'inizio della chemioterapia.



Nell'analisi della sopravvivenza si deve innanzitutto stabilire qual è l'evento terminale:

- 1) il decesso del paziente (overall survival)
- 2) il decesso legato alla malattia (disease-related survival)
- 3) il decesso legato alla malattia o la recidiva di malattia (disease-free survival)



PROBLEMI METODOLOGICI

- 1) Molti dati sono TRONCATI (CENSORED): lo studio in genere si interrompa prima che molti pazienti presentino l'evento terminale (USCITI VIVI) o alcuni pazienti possono essere persi di vista (PERSI AL FOLLOW-UP)
- 2) La distribuzione del TEMPO DI SOPRAVVIVENZA in genere presenta una forte ASIMMETRIA POSITIVA: ad esempio, molti pazienti oncologici muoiono a breve distanza dalla diagnosi o dall'intervento chirurgico, ma alcuni pazienti presentano tempi di sopravvivenza particolarmente lunghi (long survivors).

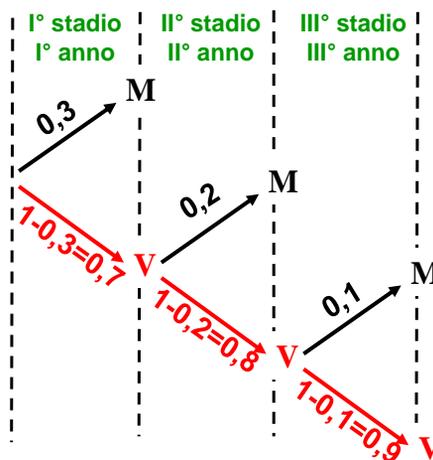
Quoziente di mortalità, $q = \frac{\text{Numero di eventi terminali}}{\text{Numero di individui esposti al rischio}}$

q non tiene conto della durata del periodo di osservazione.
 “In the long run we are all dead”.

p (probabilità di sopravvivenza) = $1 - q$ (probabilità di morte)

tasso di mortalità, $\lambda(t)$ = $\frac{\text{Numero di eventi terminali}}{\sum \text{tempi di partecipazione allo studio degli individui esposti al rischio}}$
 hazard rate

In un paziente affetto da un determinato tipo di tumore, la probabilità di morire nel I anno dalla diagnosi è del 30%, se arriva vivo alla fine del I anno la probabilità di morire nel II anno è del 20% e se arriva vivo alla fine del II anno la probabilità di morire nel III anno è del 10%.



Prob. condizionale di sopravvivenza

0,7 0,8 0,9

Prob. cumulativa di sopravvivenza

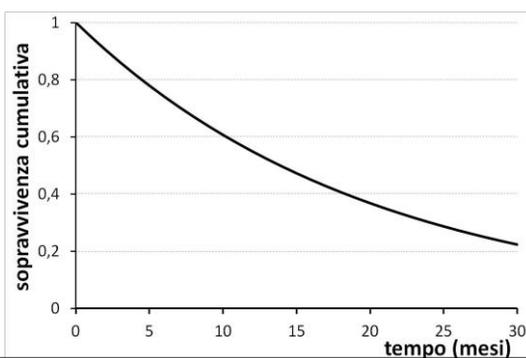
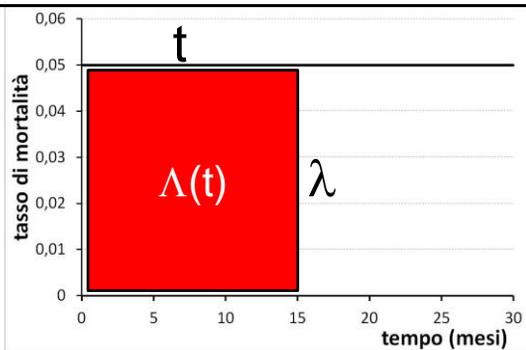
0,7 0,7*0,8
 =0,56 0,56*0,9
 =0,504

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

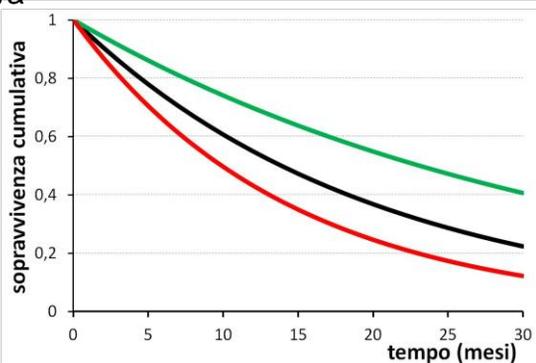
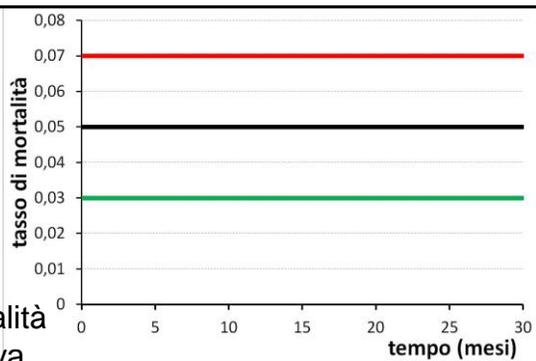
dove $\Lambda(t) = \int_0^t \lambda(u) du$ è il
cumulative hazard

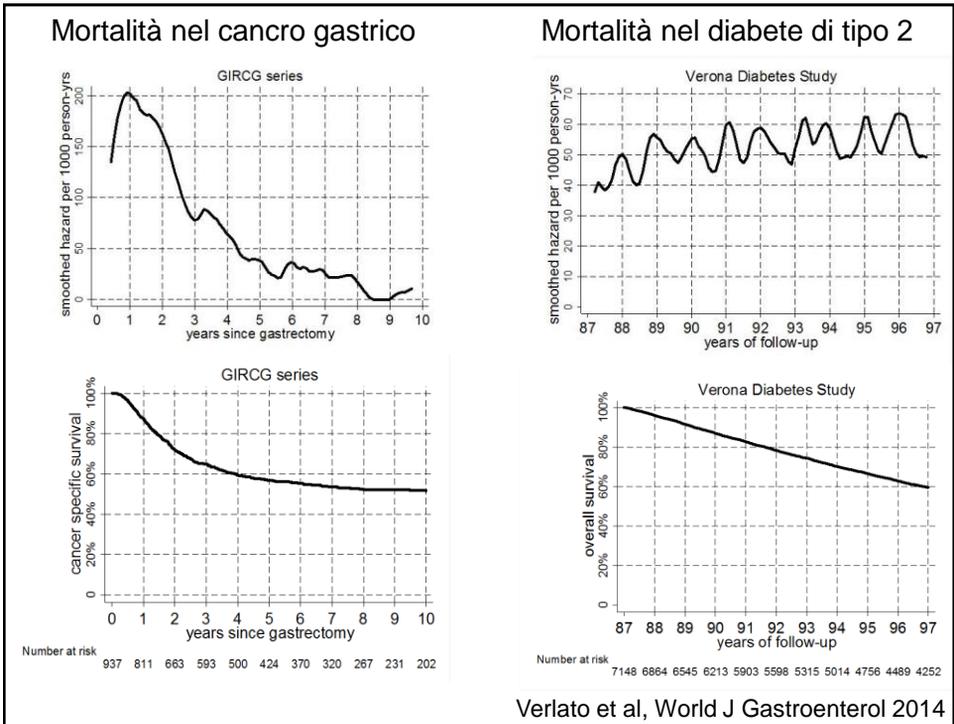
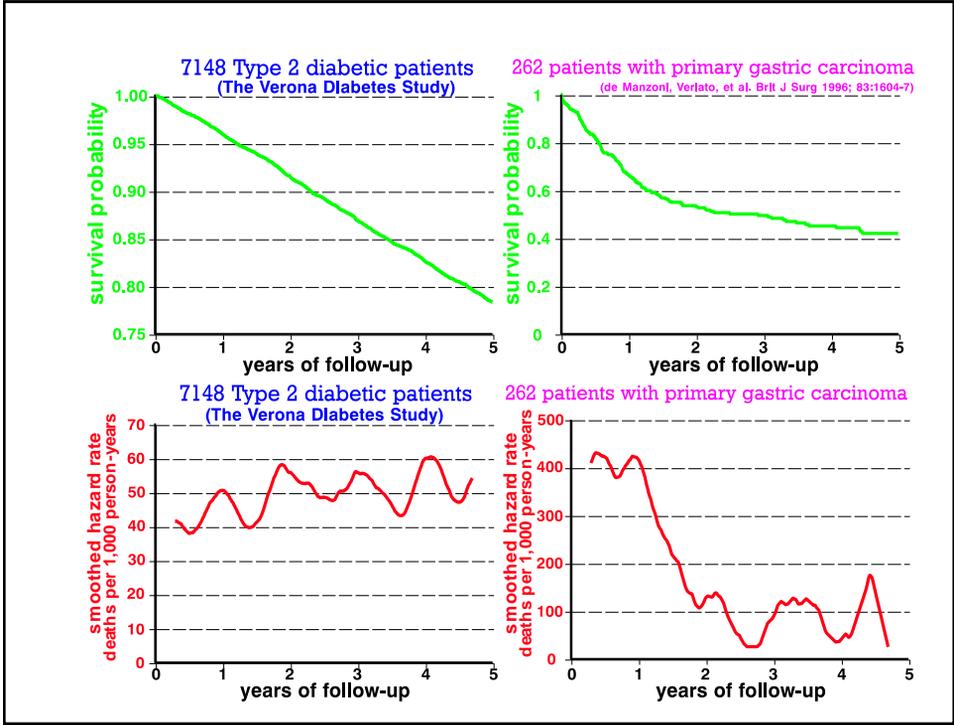
In questo caso:

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



Se aumenta il tasso di mortalità
la sopravvivenza cumulativa
cala più rapidamente





Verlato et al, World J Gastroenterol 2014

ANALISI DELLA SOPRAVVIVENZA

1) FASE DESCRITTIVA

Curve di sopravvivenza secondo il metodo del prodotto limite di Kaplan-Meier

2) FASE INFERENZIALE

Stime della sopravvivenza: metodo di Kaplan-Meier

Analisi univariabile: Confronto fra due o più curve di sopravvivenza (log-rank test)

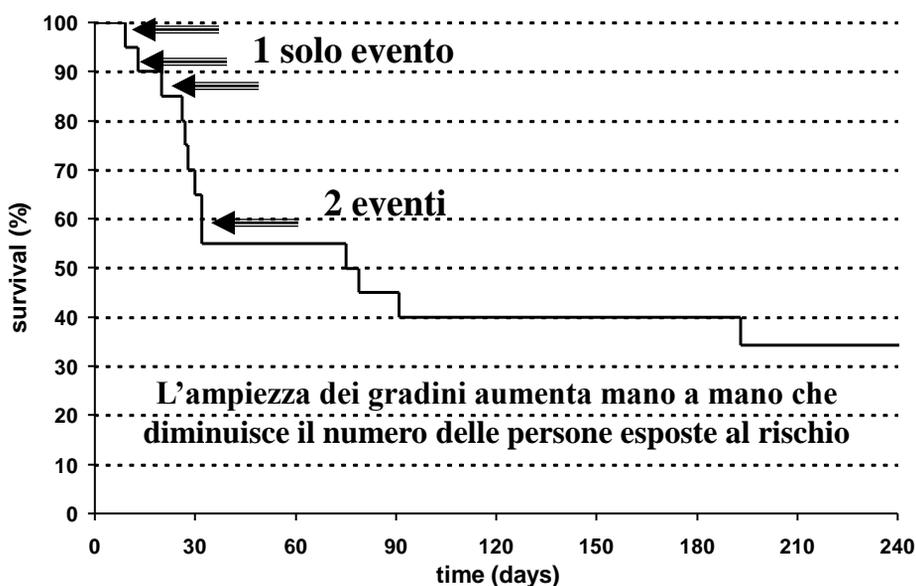
Analisi multivariabile: Valutazione simultanea del significato prognostico di più fattori di rischio (modello di Cox)

Curve di sopravvivenza secondo il metodo di Kaplan-Meier						
$q(t) =$	rischio istantaneo di morte al tempo $t =$ eventi / esposti					
$p(t) =$	probabilità condizionale di sopravvivere oltre il tempo t , dato che si è giunti vivi al tempo $t =$					$1 - q(t)$
	$\lambda(t)$ e $p(t)$ sono eventi mutuamente esclusivi ed esaustivi					
$P(t) =$	probabilità cumulativa di sopravvivere dall'inizio dello studio oltre il tempo $t =$					$p(1) * p(2) * \dots * p(t)$
	regola del prodotto di probabilità, per eventi statisticamente indipendenti					
tempo	eventi	esposti	$q(t)$	$p(t)$	$P(t)$	$P(t) \%$
0					1,0000	100,000
9	1	20	0,0500	0,9500	0,9500	95,000
13	1	19	0,0526	0,9474	0,9000	90,000
20	1	18	0,0556	0,9444	0,8500	85,000
26	1	17	0,0588	0,9412	0,8000	80,000
27	1	16	0,0625	0,9375	0,7500	75,000
28	1	15	0,0667	0,9333	0,7000	70,000
30	1	14	0,0714	0,9286	0,6500	65,000
32	2	13	0,1538	0,8462	0,5500	55,000
75	1	11	0,0909	0,9091	0,5000	50,000
79	1	10	0,1000	0,9000	0,4500	45,000
91	1	9	0,1111	0,8889	0,4000	40,000
193	1	7	0,1429	0,8571	0,3429	34,286
541	1	6	0,1667	0,8333	0,2857	28,571
1129	1	5	0,2000	0,8000	0,2286	22,857
1585	1	3	0,3333	0,6667	0,1524	15,238

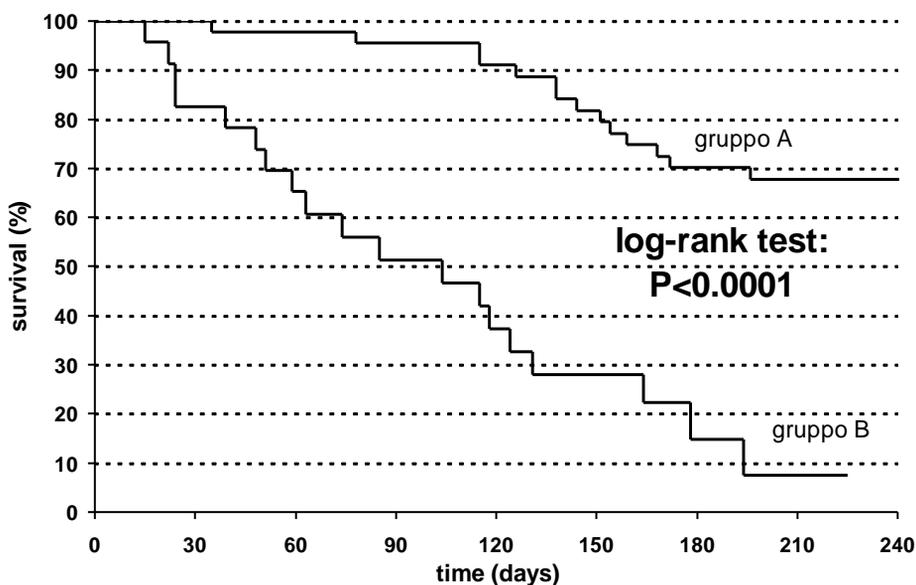
Base dati per la costruzione della curva di sopravvivenza

x	y			
tempo	P(t) %			
0	100,000	segmento		
9	100,000	orizzontale	segmento	
9	95,000		verticale	segmento
13	95,000	segmento		orizzontale
13	90,000	verticale	segmento	
20	90,000		orizzontale	segmento
20	85,000	segmento		verticale
26	85,000	orizzontale	segmento	
26	80,000		verticale	segmento
27	80,000			orizzontale
27	75,000			
28	75,000			
28	70,000			
30	70,000			
30	65,000			
32	65,000			
32	55,000			
75	55,000			
75	50,000			

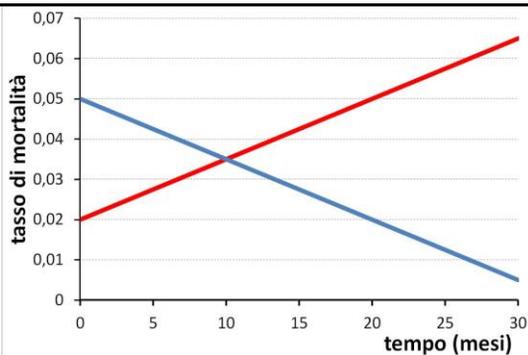
I gradini sono in corrispondenza degli eventi



Confronto fra due curve di sopravvivenza: il log-rank test

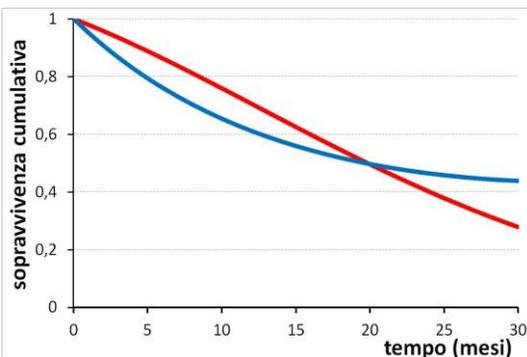


Il tasso di mortalità è maggiore nel gruppo Blu all'inizio dello studio e nel gruppo Rosso alla fine.



Anche le curve di sopravvivenza si incrociano.

Ahi ! Ahi ! Ahi !



E' una situazione che può verificarsi in campo oncologico:

- 1) i pazienti con tumore in fase iniziale vengono sottoposti ad interventi chirurgici, gravati da una percentuale di mortalità post-operatoria.
- 2) I pazienti con tumore avanzato sono spesso inoperabili, per cui presentano una prognosi a distanza molto peggiore, ma non sperimentano la mortalità post-operatoria.

Quando le curve di sopravvivenza si incrociano, il log-rank test non è più indicato.

Si può adottare il test di Wilcoxon-Breslow-Gehan, che attribuisce un peso maggiore alle fasi iniziali della curva di sopravvivenza.

Analizzare separatamente la fase iniziale e la fase tardiva del periodo di follow-up rispecchia maggiormente la realtà.

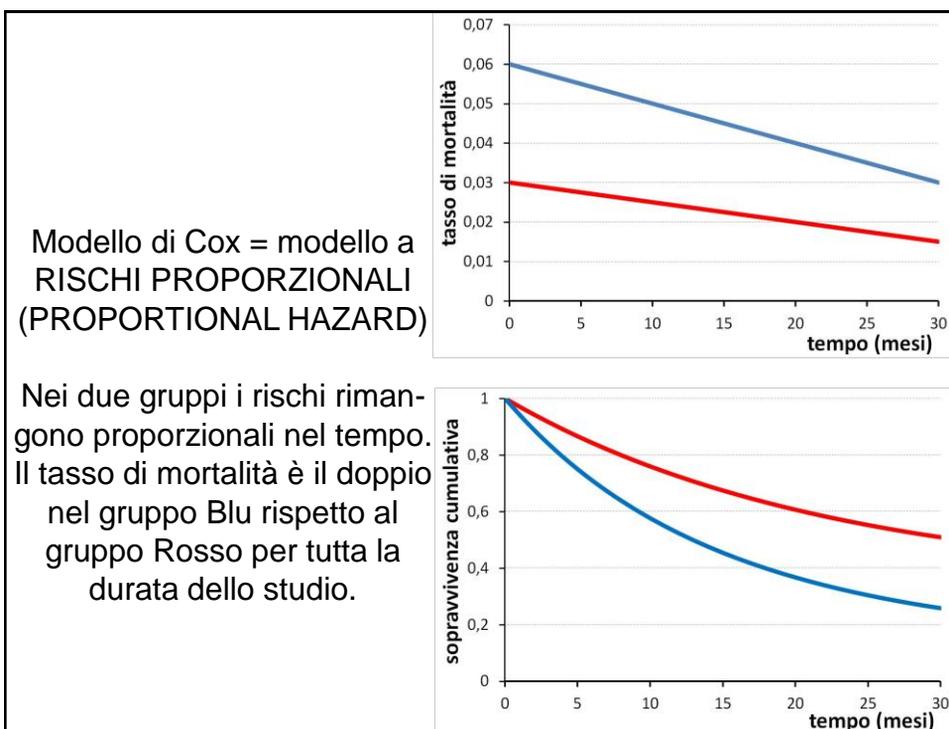
Il modello di Cox (Cox regression model) non fa alcuna assunzione a priori sull'andamento del tasso di mortalità nel tempo.

$$\lambda(t) = \lambda_0(t) * e^{-b_0 + b_1 * \text{sex} + b_2 * \text{age}}$$

funzione rischio basale
(basal hazard function)

predittore lineare: consente di
calcolare l'HR (Hazard Ratio)

Il modello di Cox è un modello a rischi proporzionali.



Per verificare se l'assunto di proporzionalità dei rischi
Si possono usare:

- 1) Metodi grafici, basati sull'andamento nel tempo di $\log [-\log(S(t))]$, equivalente a $\log[\Lambda(t)]$. L'assunzione di proporzionalità dei rischi non è violata quando le curve relative ai vari gruppi sono parallele.
- 2) Metodi statistici. Si verifica la significatività dell'interazione tra una variabile esplicativa e il $\log(\text{tempo})$

Se l'assunto di proporzionalità dei rischi è violato per una variabile esplicativa, si può utilizzare un modello di Cox stratificato per quella variabile.
In questo caso tuttavia si rinuncia a valutare il significato prognostico di quella variabile.