

Statistica Medica

Dott.ssa Marta Di Nicola

N.P.D. 3° Blocco 2° piano

0871-3554007

m.dinicola@unich.it

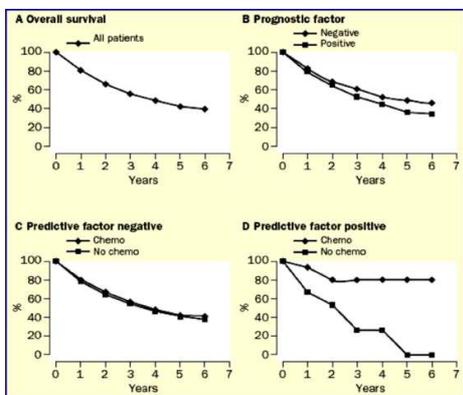
<http://www.biostatistica.unich.it>

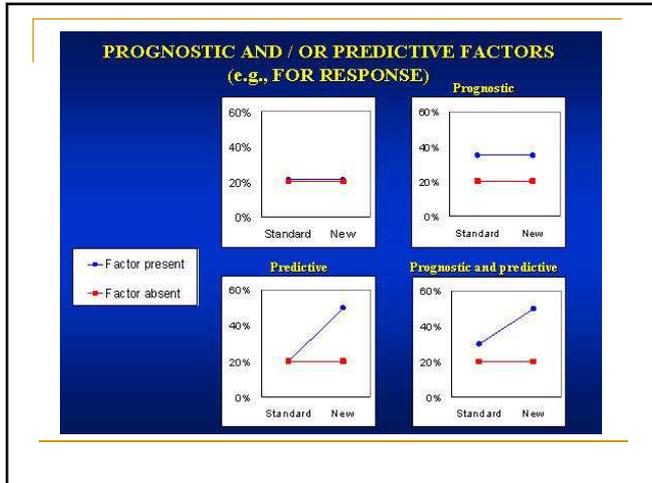
La prognosi

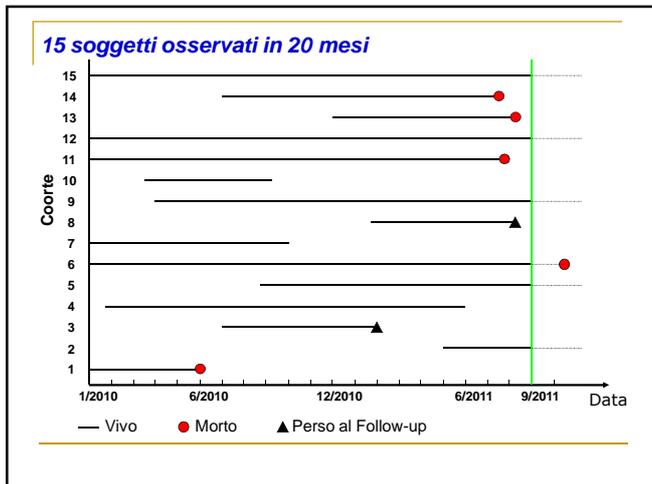
Predizione della durata, decorso ed esito di una malattia in un individuo (dunque del gruppo di individui a cui appartiene!)

Domande

- Quali fattori influenzano la prognosi?
- Qual è il peso relativo di ogni fattore?
- L'effetto di una variabile dipende dal livello di un'altra (predittività)?





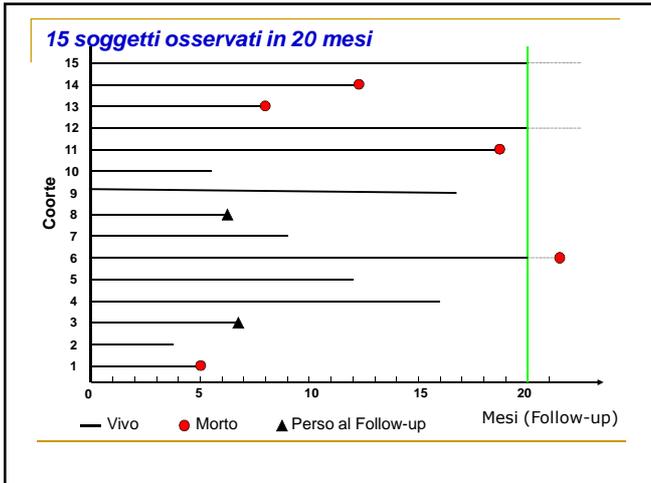


Analisi di dati dipendenti dal tempo

Una variabile di risposta dipendente dal tempo misura il tempo intercorrente fra l'inizio dell'osservazione (la diagnosi, la randomizzazione, il trattamento, la chirurgia) e il momento in cui l'evento si verifica (la morte, la recidiva).

L'obiettivo dello studio è valutare non solo se si è osservato l'esito in studio, ma anche quando esso si è verificato e, nel caso di confronto, se questo tempo è sostanzialmente diverso fra i gruppi confrontati.

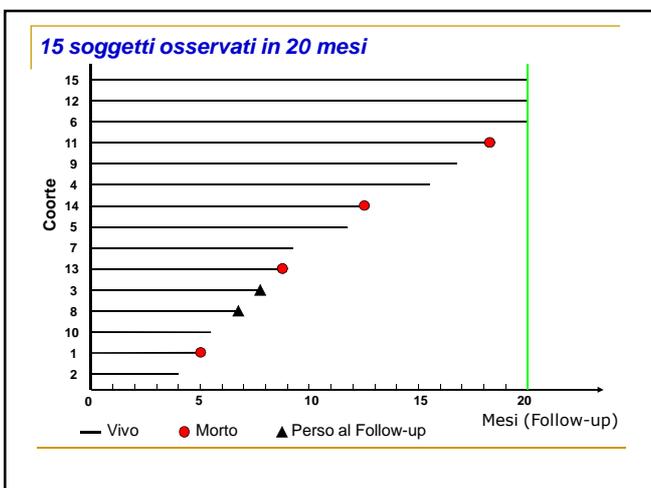
Si è verificato l'evento in studio?
 → SI → Quando?
 → NO → Fino a quando il paziente è stato osservato?



Analisi di dati dipendenti dal tempo

- I pazienti entrano nello studio in momenti differenti e, quindi, sono seguiti per periodi di tempo differenti.
- Il tempo fra l'inizio dell'osservazione e l'esito (tempo di sopravvivenza) è noto solo per i pazienti che presentano l'esito.
- Per altri pazienti l'esito in studio non si presenta (*censored*):
 - Coloro che non vanno incontro all'evento
 - Coloro che sono persi al follow-up
 - Coloro che non sono più a rischio

Per questi soggetti abbiamo un'informazione solo parziale: che il loro tempo di sopravvivenza è almeno pari al tempo di osservazione, ma non sappiamo quanto sia effettivamente.



Analisi di dati dipendenti dal tempo

Quoziente di mortalità totale

n° di eventi totali / n° casi

Nell'esempio 4/15=26.7%

LIMITI:

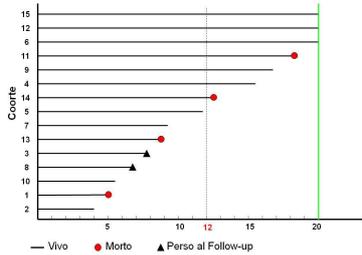
- dipende dalla durata del periodo di osservazione
- periodi di osservazione diversi per i diversi soggetti

Analisi di dati dipendenti dal tempo

Quoziente di mortalità a un dato tempo t_i

n° eventi entro t_i / n° casi con $t \geq t_i$

Nell'esempio a 12 mesi
2/7=28.6%



Analisi di dati dipendenti dal tempo

LIMITI:

- non tiene conto del tempo effettivo di osservazione di ogni soggetto
- può indurre errori sistematici
 - selezione dei pazienti da analizzare
 - selezione del tempo t_i più conveniente per l'analisi

Analisi di dati dipendenti dal tempo-SINTESI

Tempo medio di sopravvivenza

- **media aritmetica:** $\Sigma t_s / n^\circ \text{ casi}$
 - ❑ la disponibilità dei tempi di sopravvivenza di tutti i soggetti
 - ❑ distribuzione simmetrica dei tempi di sopravvivenza
- **media armonica:** $n / \Sigma (1/t_s)$
 - ❑ disponibilità dei tempi di sopravvivenza di tutti i soggetti

Tempo mediano di sopravvivenza

- ❑ disponibilità dei tempi di sopravvivenza in più del 50% dei soggetti

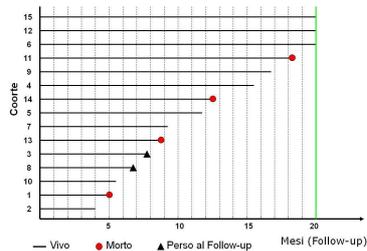
Analisi di dati dipendenti dal tempo-SINTESI

Tasso di mortalità (incidenza)

n° di eventi totali / Σ tempi di osservazione

- ❑ tasso di mortalità costante per tutto il periodo di osservazione

4 casi/186 mesi =
0.021/mese=0.25/anno

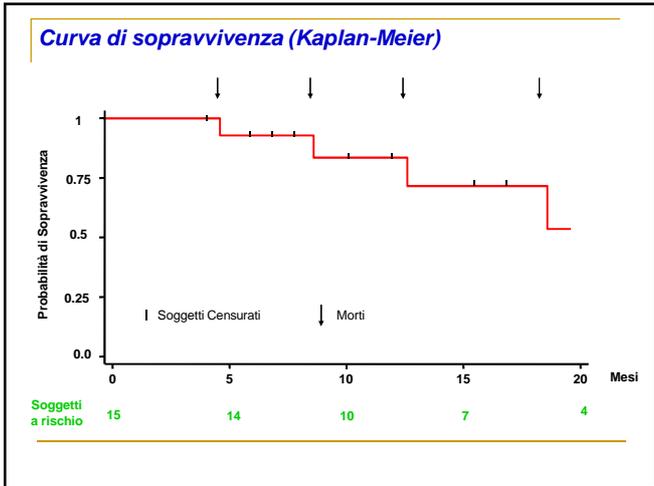


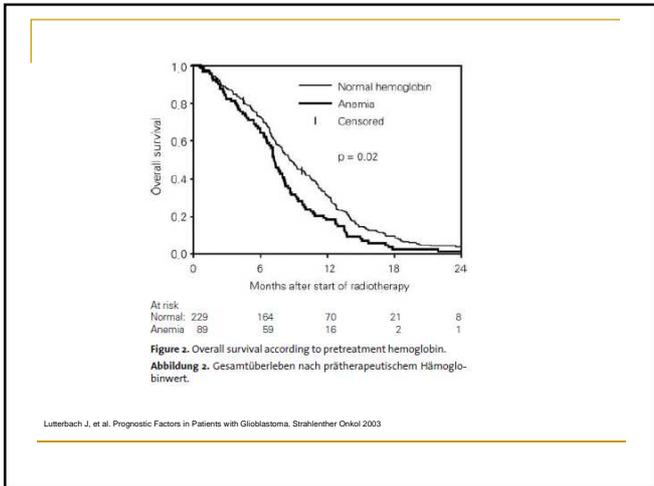
Curva di sopravvivenza

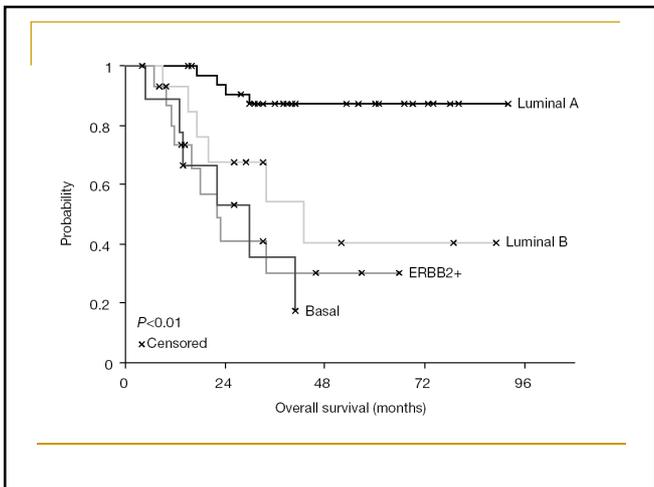
Descrive la probabilità di sopravvivere nel corso del tempo

- ❑ include **tutti** i soggetti
- ❑ valuta l'intero periodo di studio
- ❑ tiene conto del tempo effettivo di osservazione di ogni soggetto
- ❑ tiene conto di tassi di mortalità variabili nel tempo

L'idea alla base della curva di sopravvivenza è molto semplice: per sopravvivere un anno un individuo deve sopravvivere il primo giorno, il secondo, il terzo ... fino al 365-esimo giorno.







Calcolo curva di sopravvivenza (K-M)

# Pz	mesi	Esito	# a rischio	P(morte) _t	P(vivo) _t	P(vivo) _{cum}
2	4	no	15	0	1	1
1	5	si	14	0,07 (1/14)	0,93 (1-0,07)	0,93
10	6	no	13	0	1	0,93
8	7	no	12	0	1	0,93
3	8	no	11	0	1	0,93
13	9	si	10	0,10 (1/10)	0,90	0,84
7	10	no	9	0	1	0,84
5	12	no	8	0	1	0,84
14	13	si	7	0,14 (1/7)	0,86	0,71
4	16	no	6	0	1	0,71
9	17	no	5	0	1	0,71
11	19	si	4	0,25 (1/4)	0,75	0,54
6-12-15	20	no	3	0	1	0,54

La curva di sopravvivenza

- parte sempre da 1 (per definizione tutti i soggetti sono vivi al tempo 0)
- decresce nel tempo
- cambia valore solo in corrispondenza dei tempi in cui si osserva almeno un evento
- l'altezza dei gradini dipende sia dal numero di eventi che dal numero di soggetti a rischio. Anche un solo evento su pochi soggetti a rischio può modificare fortemente la stima della sopravvivenza.
- presuppone che i soggetti censurati non "nascondano" un'esperienza di sopravvivenza diversa da quella degli altri pazienti.

Come leggere una curva di sopravvivenza?

- E' importante osservare la forma della curva, più che i dettagli.
- Di solito si riportano i valori della stima della probabilità di sopravvivenza a punti rilevanti (ad esempio 1 anno)
- E' buona norma riportare il numero dei soggetti a rischio sotto la curva, in corrispondenza dei tempi
- E' buona norma "leggere" la curva sino a quando vi siano almeno 10-20 soggetti a rischio.
- E' importante conoscere la percentuale dei soggetti con dati censurati perché persi al follow-up.

Confronto tra curve di sopravvivenza

Trattamento (n=8)

Controllo (n=7)

# Pz	giorni	Esito	# Pz	giorni	Esito
1	69	si	9	17	si
2	163	si	10	39	si
3	231	no	11	56	si
4	337	no	12	98	si
5	633	si	13	152	no
6	809	no	14	252	no
7	1.046	no	15	457	si
8	1.437	no			

Confronto tra curve di sopravvivenza

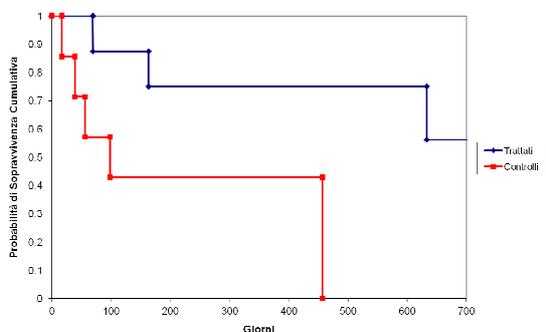
Trattati

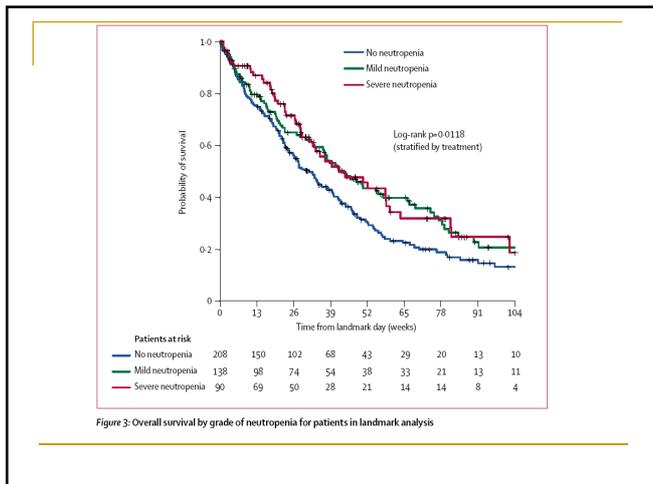
# Pz	giorni	Esito	# a rischio	P(morte) _t	P(vivo) _t	P(vivo) _{cum}
1	69	si	8	0,125	0,875	0,875
2	163	si	7	0,143	0,857	0,750
5	633	si	4	0,250	0,750	0,562

Controlli

# Pz	giorni	Esito	# a rischio	P(morte) _t	P(vivo) _t	P(vivo) _{cum}
9	17	si	7	0,143	0,857	0,857
10	39	si	6	0,167	0,833	0,714
11	56	si	5	0,200	0,800	0,571
12	98	si	4	0,250	0,750	0,428
15	457	si	1	1	0	0

Curve di sopravvivenza (KM)





Log rank test

Test statistico non parametrico utilizzato per confrontare due curve di sopravvivenza

Per ogni intervallo si costruisce una tabella 2x2

	Observed # deaths		vs.	Expected # deaths		Variance
	Dead	Alive				
Group 1	\hat{O}_1	\hat{A}_1	$R_{1j} \rightarrow$	$E_{1j} = \frac{R_{1j} C_{1j}}{n_j}$		$V_j = \frac{\hat{A}_1 \hat{R}_{2j} C_{1j} C_{2j}}{n_j^2 (n_j - 1)}$
Group 2	\hat{O}_2	\hat{A}_2	$R_{2j} \rightarrow$	$E_{2j} = \frac{\hat{R}_{2j} C_{2j}}{n_j}$		
	C_{1j}	C_{2j}		n_j		

Sommando per tutti gli intervalli si ottiene

Observed total deaths	Expected total deaths	Total Variance
Group 1: $O_1 = \sum \hat{O}_1$	Group 1: $E_1 = \sum E_{1j}$	$V = \sum V_j$
Group 2: $O_2 = \sum \hat{O}_2$	Group 2: $E_2 = \sum E_{2j}$	

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} \sim \chi_1^2$$

Cox proportional hazard model

La misura dell'efficacia con dati dipendenti dal tempo

RR, RRR, ARR, NNT possono essere applicate anche alle curve di sopravvivenza, una volta che si fissi un tempo (t), il cui limite è di essere scelto in modo arbitrario.

L' **Hazard ratio** (HR), o rapporto degli hazard, è una misura relativa e può essere interpretato come il rapporto fra i tassi medi (o velocità medie) di mortalità nei gruppi confrontati.

Una ulteriore misura assoluta per confrontare le curve di sopravvivenza è la **differenza fra le mediane di sopravvivenza**. Essa indica di quanto si è modificato il tempo a cui si osserva il 50% di morti fra i trattati rispetto ai controlli.

La misura dell'efficacia con dati dipendenti dal tempo

Rischio relativo $RR = F_i(t)/F_c(t)$

Riduzione relativa del rischio $RRR = 1 - RR$

'Odds Ratio' $OR = F_i(t)S_c(t)/F_c(t)S_i(t)$

Riduzione assoluta del rischio $ARR = F_c(t) - F_i(t)$

'Number needed to be treated' $NNT = 1/ARR$

'Hazard Ratio' (HR) $HR = h_i(t)/h_c(t)$
