

L'analisi di sopravvivenza con il metodo di Kaplan-Meier

G. Tripepi, F. Catalano

CNR-IBIM, Consiglio Nazionale delle Ricerche, Istituto di Biomedicina Epidemiologia Clinica e Fisiopatologia delle Malattie Renali e dell'Ipertensione Arteriosa, Reggio Calabria

Kaplan-Meier analysis

Kaplan-Meier analysis is largely used in nephrology to estimate a population survival curve from a sample. If patients are followed-up until death, the survival curve can be estimated simply by computing the fraction survival at each time point. Kaplan-Meier analysis allows the estimation of survival over time, even when patients drop out or are studied for different lengths of time. In Kaplan-Meier analysis, the comparison between two survival curves is made by the log-rank test. (G Ital Nefrol 2004; 21: 540-6)

KEY WORDS: Survival analysis, Kaplan-Meier analysis, Survival time, Conditional probability, Observed risk

PAROLE CHIAVE: Analisi di sopravvivenza, Analisi di Kaplan-Meier, Tempo di sopravvivenza, Probabilità condizionata, Rischio osservato

Introduzione

L'importanza della statistica nella ricerca bio-medica sta nel fatto che essa consente di trarre delle conclusioni o fare inferenze facendo riferimento ad un campione piuttosto che a tutte le unità costituenti una popolazione. Una volta estratto un campione da una popolazione è possibile testare su di esso una o più ipotesi attraverso l'uso congiunto della statistica descrittiva e della statistica inferenziale. La statistica descrittiva consente di descrivere il campione. Per esempio, un campione di persone può essere descritto sulla base delle sue caratteristiche come l'età, il sesso, il fumo e la statura. Questi fattori possono essere a loro volta riassunti, generalmente, attraverso la media o la mediana¹ oppure tramite misure di frequenza (proporzione di maschi, percentuale di fumatori, etc). L'analisi di sopravvivenza è parte della statistica inferenziale e la sua peculiarità consiste nel mettere in rapporto un certo esito o evento con il fattore tempo. Ad esempio, se volessimo studiare la sopravvivenza a due anni di un campione di persone (coorte dello studio) dovremmo seguire ciascuno di loro per 24 mesi. Alla fine di questo periodo di osservazione ciascun paziente sarebbe descritto da una coppia di valori che indichiamo generalmente con (c, t). Dove "c" è la condizione (0=non morto, 1=morto) e "t" è la durata dell'osservazione. La caratteristica principale delle coorti è la loro dinamicità:

infatti seguire una coorte di persone per un tempo "t" implica la possibilità che la sua numerosità vari nel tempo in ragione del verificarsi o meno dell'evento di interesse (per esempio l'evento morte).

Prima di addentrarci nei dettagli dell'analisi di Kaplan-Meier è utile chiarire tre concetti di base, ovvero: tempo di sopravvivenza, probabilità condizionata e rischio osservato.

Il tempo di sopravvivenza

Il termine "tempo di sopravvivenza" va usato in senso estensivo perché si applica anche a eventi diversi dalla morte. L'analisi di sopravvivenza riguarda infatti tutti quegli studi in cui si vuole analizzare l'incidenza² di un determinato evento in un certo arco temporale (studi di coorte). Perciò il tempo di sopravvivenza assume significati diversi in relazione al tipo di evento a cui è interessato il ricercato-

¹ La mediana è quella osservazione il cui valore è tale da dividere in due gruppi di eguale numerosità un determinato campione ovvero è quella osservazione il cui valore è tale che metà delle osservazioni siano maggiori e metà inferiori ad essa.

² L'incidenza di un determinato evento in una coorte di pazienti corrisponde al numero di pazienti che sviluppano l'evento in un determinato arco temporale.

re. Pertanto il tempo di sopravvivenza può essere il tempo che intercorre tra l'inizio dello studio e la morte, l'incidenza di un evento cardiovascolare (come l'infarto o l'ictus), il tempo di insorgenza di una patologia (per esempio le peritoniti in una coorte di pazienti in CAPD), il tempo dopo il quale si osserva un aumento significativo dei livelli plasmatici di un certo marker biochimico (per esempio il raddoppio della creatinina o della proteinuria in una coorte di pazienti con insufficienza renale lieve) o il tempo in cui si verifica un episodio di rigetto renale (in una coorte di pazienti portatori di trapianto di rene). Ovviamente il tempo di sopravvivenza, per i pazienti senza l'evento di interesse corrisponde a quello compreso tra l'ingresso nello studio e la fine dell'osservazione.

La probabilità condizionata

Il rapporto tra l'analisi di sopravvivenza e la probabilità condizionata è semplice ed intuitivo come del resto è chiara l'identità tra rischio e probabilità. Infatti la probabilità che ha un paziente di sopravvivere dopo tre giorni dall'ingresso in uno studio (probabilità che indichiamo generalmente con P) è condizionata dal fatto che il paziente sia sopravvissuto nei due giorni precedenti. Questa probabilità viene anche definita probabilità cumulativa o sopravvivenza cumulativa. Se indichiamo con p_1 la probabilità che ha il paziente di sopravvivere il primo giorno, con p_2 la probabilità di sopravvivere il secondo giorno e con p_3 la probabilità di sopravvivere il terzo giorno, la sopravvivenza cumulativa è data dal prodotto di queste singole probabilità ($P=p_1 \cdot p_2 \cdot p_3$) (regola moltiplicatoria della probabilità). Per cui se la probabilità di sopravvivere il primo giorno è 0.80, quella di sopravvivere il secondo giorno è il 0.68 e quella di sopravvivere il terzo giorno è 0.55 la probabilità cumulativa è $0.80 \cdot 0.68 \cdot 0.55$ ovvero 0.30 (cioè il 30%). Un ulteriore esempio che ben si addice a spiegare la probabilità condizionata è il lancio ripetuto di una moneta. Se la moneta è perfettamente bilanciata, la probabilità che venga testa è 0.5 (il 50%) per ogni lancio. Se volessimo sapere qual è la probabilità che venga testa in tre lanci successivi (che rappresenta la "condizione" dell'esperimento) il calcolo da fare è: $0.5 \cdot 0.5 \cdot 0.5 = 0.125$ (ovvero il 12.5%).

Il rischio osservato

Il rischio osservato è il rischio misurato osservando la realtà. Per comprendere meglio il concetto di rischio osservato supponiamo di avere a che fare con 11 pazienti ipertesi di cui 4 con severa ipertrofia ventricolare sinistra (IVS) e 7 senza IVS. Ipotizziamo di aver seguito questi pazienti per 1 anno. Il nostro obiettivo è quello di misurare il rischio relativo per un dato evento cardiovascolare (per esempio

		Fattore di rischio: Ipertrofia ventricolare sinistra	
		Presente	Assente
Evento	No	1	6
	Si	3	1
Totale		4	7

Fig. 1

l'infarto del miocardio) associato alla presenza dell'IVS. Costruiamo una tabella 2 x 2. Sulle colonne riportiamo il fattore di rischio (presente/assente) e sulle righe l'evento cardiovascolare (si/no) (Fig. 1). Alla fine del periodo 3 dei 4 pazienti con marcata IVS hanno avuto 1 evento cardiovascolare mentre solo 1 dei 7 pazienti senza IVS ha avuto un evento. Per calcolare il rischio relativo di eventi cardiovascolari a cui sono esposti i pazienti del gruppo con IVS rispetto al gruppo senza IVS basta fare: $(3/4)/(1/7)$: ovvero 5.4. Ciò vuol dire che i pazienti con IVS hanno un rischio osservato di eventi cardiovascolari che è 5.4 volte maggiore rispetto ai pazienti senza IVS. Lo strumento di analisi statistica che ci consente di costruire le curve di sopravvivenza (ovvero il grafico della relazione esistente tra la probabilità di sopravvivere e il tempo di osservazione) e di misurare il rischio osservato sono le curve di sopravvivenza di KAPLAN-MEIER.

Test di verifica

1) Cosa s'intende per tempo di sopravvivenza?

- È il tempo che intercorre tra l'ingresso nello studio e il solo evento morte
- È esclusivamente il tempo che intercorre tra l'ingresso nello studio e la data in cui il paziente viene censurato
- È un termine che va usato in senso estensivo perché si applica anche a eventi diversi dalla morte
- È il tempo trascorso tra l'ingresso e l'uscita dallo studio e va calcolato solo per i pazienti sopravvissuti.

2) Cosa s'intende per probabilità condizionata?

- È la probabilità che un evento ha di verificarsi in un intervallo di tempo tenendo conto della somma delle probabilità che l'evento ha di verificarsi negli intervalli di tempo precedenti
- È la probabilità che un evento ha di verificarsi in un intervallo di tempo tenendo conto delle probabilità che l'evento ha di verificarsi negli intervalli di tempo pre-

- cedenti. Si ottiene dal prodotto delle probabilità dell'evento nei singoli intervalli di tempo
- c. È la probabilità che un evento ha di verificarsi durante tutto l'intero periodo di osservazione e coincide con la probabilità iniziale di morte
 - d. È la probabilità assoluta che un evento ha di verificarsi in un successivo intervallo di tempo indipendentemente dalla probabilità che l'evento ha di verificarsi negli intervalli di tempo precedenti.

3) Cosa s'intende per rischio relativo?

- a. È il rapporto tra la proporzione di pazienti che presentano l'evento nel gruppo esposto al fattore di rischio e la proporzione di pazienti che non presentano l'evento nel gruppo non esposto al fattore di rischio
- b. È il rapporto tra il numero di pazienti che presentano l'evento nel gruppo esposto al fattore di rischio e il numero dei pazienti che non presentano l'evento e che sono esposti al medesimo fattore di rischio
- c. È il rapporto tra la proporzione di pazienti che presentano l'evento nel gruppo esposto al fattore di rischio e la proporzione di pazienti che presentano l'evento nel gruppo non esposto al fattore di rischio
- d. È il rapporto tra il numero dei pazienti che non presentano l'evento nel gruppo esposto al fattore di rischio e il numero dei pazienti che non presentano l'evento nel gruppo non esposto al fattore di rischio.

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e sul prossimo numero del Giornale cartaceo

L'analisi di Kaplan-Meier

Come costruire una curva di sopravvivenza

Prendiamo in considerazione 10 pazienti con severa IVS e supponiamo di seguirli per 240 giorni. Il nostro obiettivo è quello di costruire la curva di sopravvivenza in relazione all'incidenza dell'infarto del miocardio fatale. Durante la

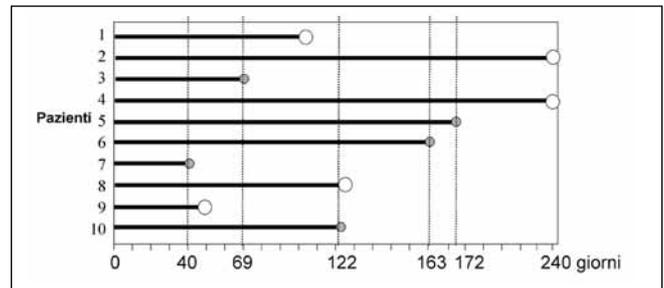


Fig. 2

descrizione del metodo chiameremo “censurati” tutti i pazienti che non hanno l'infarto del miocardio durante il periodo di osservazione ovvero coloro che sopravvivono fino alla fine dell'osservazione oppure che escono dall'osservazione prima del termine dello studio per ragioni diverse dall'infarto del miocardio (per esempio pazienti persi al follow-up, trasferiti o morti per cause diverse da quella d'interesse). È importante sottolineare che i pazienti censurati rimangono nell'analisi fino al momento in cui sono disponibili dati certi sul loro stato di salute e la loro presenza, come vedremo meglio in seguito, è di fondamentale importanza analitica. Vediamo in dettaglio le diverse situazioni che si sono presentate nel nostro ipotetico esempio (Fig. 2). I cerchietti vuoti indicano i censurati mentre i cerchietti grigi i pazienti con l'evento. Il paziente n. 1 viene censurato dopo 100 giorni perché perso al follow-up, il paziente n. 2 arriva vivo alla fine dell'osservazione, il paziente n. 3 ha l'infarto dopo 59 giorni, il paziente n. 4 arriva vivo alla fine dell'osservazione, il paziente n. 5 ha l'infarto dopo 172 giorni, il paziente n. 6 ha l'infarto dopo 163 giorni, il paziente n. 7 ha l'infarto dopo 40 giorni, il paziente n. 8 viene censurato dopo 125 giorni perché trasferito in un'altra città, il paziente n. 9 viene censurato dopo 50 giorni perché morto per neoplasia e il paziente n. 10 ha l'infarto dopo 122 giorni. Il passo successivo è quello di dividere il periodo di tempo (ovvero i 240 giorni) in intervalli. Nell'analisi di Kaplan-Meier il numero degli intervalli è dettato dai tempi in cui ha luogo l'evento di interesse. Nel nostro caso avremo complessivamente 6 intervalli, uno per ogni evento che si è verificato (Fig. 2, linee verticali tratteggiate). Il

TABELLA I

Intervallo	Giorni	A rischio nel periodo	Eventi nel periodo	Censurati nel periodo	Sopravvivenza del periodo	Sopravvivenza cumulativa
1	0-40	10	1	0	0.900	0.900
2	41-69	9	1	1	0.890	0.801
3	70-122	7	1	1	0.857	0.684
4	123-163	5	1	1	0.800	0.547
5	164-172	3	1	0	0.666	0.364
6	173-240	2	0	2	1.000	0.364

primo intervallo è 0-40 giorni, il secondo è 41-69, il terzo è 70-122, il quarto è 123-163, il quinto è 164-172, il sesto è 173-240. Riportiamo adesso in forma tabellare i dati (Tab. I). Partiamo dal primo intervallo (0-40). In questo intervallo il numero dei pazienti a rischio è 10. I pazienti a rischio in un intervallo di tempo t_1 sono i pazienti che possono potenzialmente avere l'evento nell'intervallo di tempo t_1 ovvero quelli sopravvissuti e non censurati nell'intervallo di tempo precedente più i pazienti censurati nell'intervallo di tempo t_1 . Ovviamente, all'inizio dell'osservazione i pazienti a rischio sono 10. Nel primo intervallo si verifica 1 evento e non vi sono censurati. Nel secondo intervallo (41-69 giorni) i pazienti a rischio sono 9 (1 ha avuto l'evento nell'intervallo precedente e pertanto ha terminato lo studio). Nel secondo intervallo si è verificato 1 evento e 1 paziente è stato censurato. Nel terzo intervallo (60-122 giorni) i pazienti a rischio sono 7 (in quanto nell'intervallo precedente 1 paziente ha avuto l'evento ed 1 è stato censurato). Nell'intervallo 60-122 giorni 1 paziente ha avuto 1 evento e 1 paziente è stato censurato. Nel quarto intervallo (123-163 giorni) i pazienti a rischio sono 5 (in quanto nell'intervallo precedente 1 paziente ha avuto l'evento ed 1 è stato censurato). Nel quinto intervallo (164-172 giorni) i pazienti a rischio sono 3 (in quanto nell'intervallo precedente 1 paziente ha avuto l'evento ed 1 è stato censurato). Nell'intervallo (164-172 giorni) 1 paziente ha avuto l'evento e nessun è stato censurato. Nel sesto intervallo i pazienti a rischio sono 2 (in quanto nell'intervallo precedente 1 paziente ha avuto 1 evento ma nessun è stato censurato). In questo ultimo intervallo non si è verificato nessun evento e 2 pazienti sono stati censurati. Vediamo adesso come calcolare la sopravvivenza in ciascun periodo e facciamo sempre riferimento alla Figura 2. La sopravvivenza nel periodo è la percentuale di pazienti vivi in quel periodo. Perciò, per calcolare la sopravvivenza del periodo si utilizza la formula: $[1 - \frac{\text{rapporto tra il numero di eventi nell'intervallo di tempo "t" e il numero di persone a rischio nell'intervallo di tempo "t"}]{}$. In Tabella I, sesta colonna, è riportata la sopravvivenza per ciascun periodo calcolata con questa formula. Passiamo adesso al calcolo della sopravvivenza cumulativa, intervallo per intervallo. Come già sappiamo la probabilità di sopravvivere in un certo intervallo di tempo è "condizionata" dalla probabilità di sopravvivere nell'intervallo di tempo precedente (per esempio un paziente per giungere vivo al terzo intervallo di tempo vuol dire che sia sopravvissuto anche nel secondo e nel primo intervallo). Per calcolare la probabilità cumulativa o probabilità condizionata basta fare il prodotto delle probabilità di sopravvivenza di ciascuno degli n intervalli fino all'intervallo *ennesimo* (ovvero applicare la regola moltiplicativa della probabilità). In base a quanto detto la probabilità cumulativa nel primo intervallo coincide con la probabilità di sopravvivenza (ovvero 0.90); nel secondo intervallo la sopravvivenza cumulativa è 0.80 [ovvero probabilità di sopravvivere nel primo intervallo (0.90) per la probabilità di sopravvivere nel secondo inter-

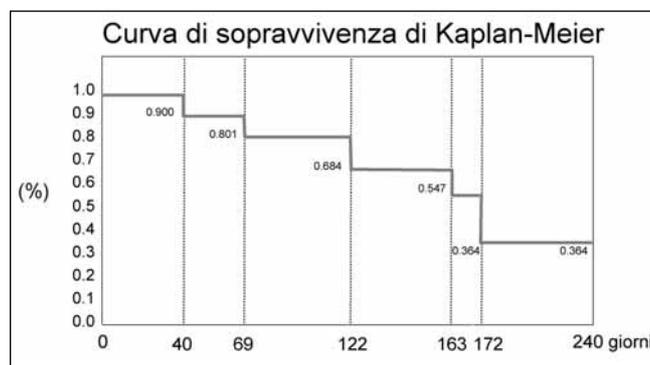


Fig. 3

vallo (0.89); $0.90 \cdot 0.89 = 0.80$]; nel terzo intervallo è 0.68 (ovvero $0.90 \cdot 0.89 \cdot 0.857$) e così via. Con questi dati a nostra disposizione e gli intervalli di tempo in giorni è possibile costruire la curva di sopravvivenza (Fig. 3). Basta riportare su di un sistema di assi cartesiani sull'asse delle ordinate la sopravvivenza cumulativa (Tab. I - settima colonna) e sull'asse delle ascisse l'intervallo di tempo in giorni.

Come confrontare due curve di sopravvivenza

Abbiamo finora visto come si costruisce una curva di sopravvivenza. Vediamo adesso come confrontare tra loro due curve di sopravvivenza. Ciò è particolarmente importante quando, per esempio, vogliamo confrontare la sopravvivenza di due gruppi di pazienti: uno esposto ed uno non esposto ad un certo fattore di rischio (per esempio pazienti con e senza ipertrofia ventricolare sinistra, fumatori e non fumatori, ipertesi e normotesi e così via). Il test da utilizzare è il **log rank test**. Nonostante il nome di questo test alluda ai logaritmi e ai ranghi esso non ha nulla a che fare né con i logaritmi né con i ranghi. Per capire il procedimento alla base del test facciamo un esempio pratico. Supponiamo di avere a che fare con due gruppi di pazienti costituiti entrambi da 6 soggetti e di voler confrontare la sopravvivenza del gruppo esposto al fattore di rischio (Gruppo 1) con quella del gruppo non esposto (Gruppo 2). L'evento che prendiamo in considerazione è l'infarto del miocardio fatale. Il nostro obiettivo è verificare se la differenza di sopravvivenza tra i due gruppi è o meno statisticamente significativa ovvero se la sopravvivenza è più bassa nel gruppo con il fattore di rischio rispetto al gruppo senza il fattore di rischio. Vediamo come procedere. Per prima cosa, elenchiamo per entrambi i gruppi, paziente per paziente, il tempo in giorni in cui ha avuto luogo l'evento o la censura (i pazienti censurati li indicheremo con l'asterisco).

Durata dell'osservazione: 100 giorni.

Gruppo 1: 23, 43, 38*, 61, 66*, 83 (giorni);

Gruppo 2: 25, 26, 36, 60*, 71, 78 (giorni).

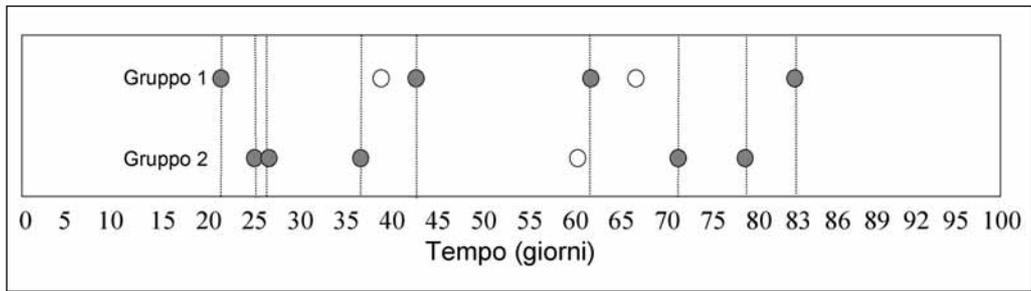


Fig. 4

TABELLA II

Intervallo	A rischio Gruppo 1	A rischio Gruppo 2	Totale a rischio (Gruppo 1+Gruppo II)	Censurati	Eventi Gruppo 1	Mortalità attesa Gruppo 1	Eventi Gruppo 2	Mortalità attesa Gruppo 2
1-23	6	6	12	0	1	6/12=0.50	0	6/12=0.50
24-25	5	6	11	0	0	5/11=0.454	1	6/11=0.546
26-26	5	5	10	0	0	5/10=0.50	1	5/10=0.50
27-36	5	4	9	0	0	5/9=0.556	1	4/9=0.444
37-43	5	3	8	1	1	5/8=0.625	0	3/8=0.375
44-61	3	3	6	1	1	3/6=0.50	0	3/6=0.50
62-71	2	2	4	1	0	2/4=0.50	1	2/4=0.50
72-78	1	1	2	0	0	1/2=0.50	1	1/2=0.50
79-83	1	0	1	0	1	1/1=1.00	0	0/1=0.00
					4 morti	5.13 morti	5 morti	3.87 morti

In questo esempio abbiamo solo due categorie di soggetti: pazienti con l'evento e censurati. Nessuno è giunto con lo status di "sopravvissuto" fino alla fine dell'osservazione (ovvero 100 giorni). Per rendere ancora più chiara la procedura, riportiamo i dati su un grafico (i cerchietti grigi indicano i pazienti con l'evento e quelli bianchi i censurati) (Fig. 4). Come già sappiamo, nell'analisi di Kaplan-Meier il numero degli intervalli è dato dagli eventi. In questo caso, poiché i gruppi sono due, la suddivisione in intervalli deve tenere conto di ciò che accade in entrambi i gruppi. Perciò, in questo caso avremo 9 intervalli. Riportiamo i dati in una tabella (Tab. II) che riporta l'intervallo di tempo, i pazienti a rischio nel Gruppo 1, i pazienti a rischio nel Gruppo 2, il totale dei pazienti a rischio, i censurati, gli eventi nel gruppo 1 e la mortalità attesa nel Gruppo 1, gli eventi nel Gruppo 2 e la mortalità attesa nel gruppo 2 (chiariremo in seguito cosa s'intende per mortalità attesa). Nel primo intervallo i pazienti a rischio nel Gruppo 1 sono 6, nel gruppo 2 sono 6, perciò il numero totale di pazienti a rischio è 12. In questo primo intervallo non vi sono censurati. È importante sottolineare che il numero dei censurati è quello totale dell'intervallo, ovvero quelli del Gruppo 1 più quelli del Gruppo 2. Passiamo adesso agli eventi nel Gruppo I (indicheremo con zero l'assenza dell'evento e con 1 l'evento). Nell'intervallo 1-23 giorni abbiamo un evento

nel Gruppo 1. Passiamo agli eventi nel Gruppo 2. Nel Gruppo 2, nell'intervallo 1-23 giorni non vi è nessun evento perciò 0. Passiamo al secondo intervallo (24-25 giorni). I pazienti a rischio nel Gruppo 1 sono 5 (1 ha avuto l'evento nell'intervallo precedente). I pazienti a rischio nel Gruppo 2 sono 6 in quanto nell'intervallo precedente non abbiamo avuto nessun evento e nessun censurato. Il totale dei pazienti a rischio è 11. I censurati sono 0. Passiamo agli eventi nel Gruppo 1. In questo intervallo gli eventi sono 0. Passiamo agli eventi nel Gruppo 2. In questo caso abbiamo 1 evento. Passiamo adesso all'intervallo (26-26 giorni). I pazienti a rischio nel Gruppo 1 sono 5 (ricordiamoci che abbiamo avuto un evento nell'intervallo 1-23 giorni nel gruppo 1). I pazienti a rischio nel Gruppo 2 sono 5 (ricordiamoci che, nel Gruppo 2, abbiamo avuto un evento nell'intervallo 24-25 giorni). In totale il numero dei pazienti a rischio nel terzo intervallo è 10. Per completare l'analisi basta ripetere i suddetti calcoli in maniera iterativa fino al nono intervallo. Alla fine avremo 4 morti per il Gruppo 1 e 5 morti nel Gruppo 2. Passiamo al calcolo della mortalità attesa. Facciamo l'esempio solo per il Gruppo 1 in quanto analogo ragionamento può essere fatto per il Gruppo 2. Per mortalità attesa nel Gruppo 1 s'intende il massimo contributo in termini di mortalità che il Gruppo 1 può dare nell'ambito del totale dei pazienti a rischio (ovvero la somma

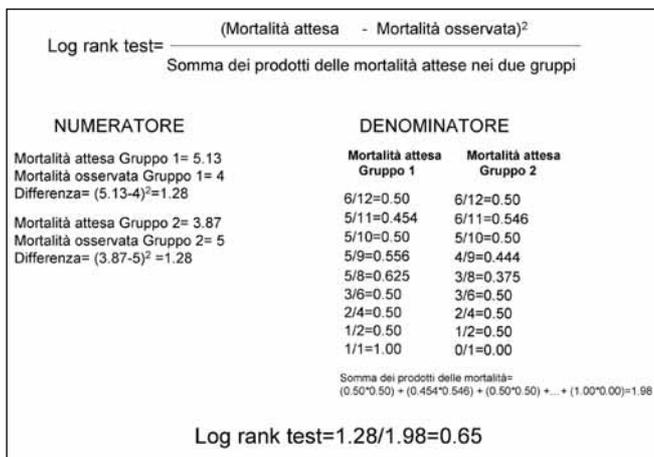


Fig. 5

dei pazienti del Gruppo 1 + Gruppo 2). Il calcolo va fatto intervallo per intervallo. È ovvio che quando i due gruppi hanno la stessa numerosità la mortalità attesa si distribuisce in eguale misura nei due gruppi (0.5 e 0.5). Sulla base di quanto detto nel primo intervallo (1-23) per quanto attiene al Gruppo 1 la mortalità attesa è 6/12=0.50. Nel Gruppo 2 è sempre 6/12=0.50. Nel secondo intervallo per quanto riguarda il Gruppo 1 è 5/11=0.454; per quanto riguarda il Gruppo 2 è 6/11=0.546. Nel terzo intervallo per quanto riguarda il Gruppo 1 è 5/10=0.50; per quanto riguarda il Gruppo 2 è 5/10=0.50. Nel quarto intervallo per quanto riguarda il Gruppo 1 è 5/9=0.556; per quanto riguarda il Gruppo 2 è 4/9=0.444. E così via. Il totale della mortalità attesa nel Gruppo 1 è 5.13; mentre il totale della mortalità attesa nel Gruppo 2 è 3.87.

Per calcolare il *log rank test* (Fig. 5) si usa la formula:

$$\frac{(Mortalità\ attesa - Mortalità\ osservata)^2}{Somma\ dei\ prodotti\ delle\ mortalità\ attese\ nei\ due\ gruppi}$$

Somma dei prodotti delle mortalità attese nei due gruppi

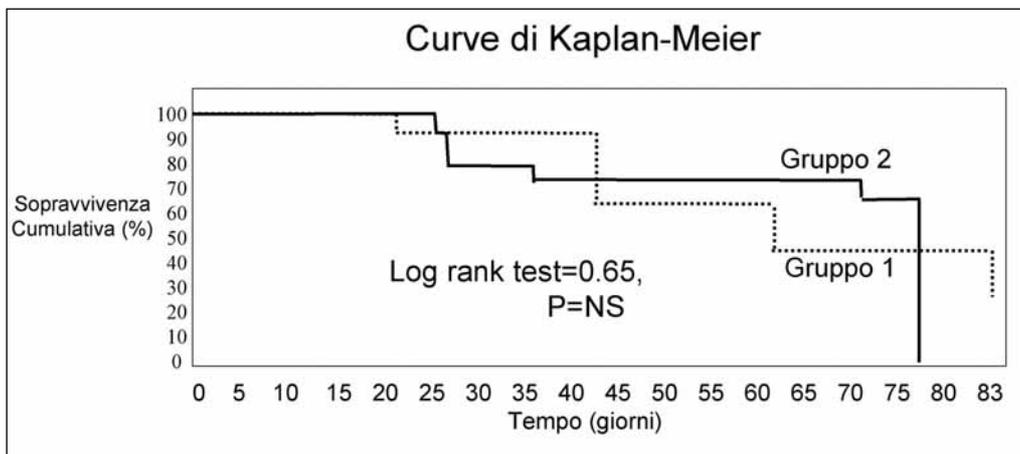


Fig. 6

Il numeratore può essere calcolato indifferentemente utilizzando i dati del Gruppo 1 o del Gruppo 2 (il risultato è infatti identico) (Fig. 5). Il denominatore si calcola moltiplicando, intervallo per intervallo, le mortalità attese nei due gruppi e sommandoli fra loro. In questo caso il log rank test è uguale a 0.65. Il risultato ottenuto va quindi tradotto in una probabilità consultando apposite tavole (quelle relative alla distribuzione del χ^2). Affinché un log rank test sia statisticamente significativo ($P < 0.05$) è necessario che abbia un valore di almeno 3.84. In Figura 6 è riportato il risultato finale dell'analisi di Kaplan-Meier relativa ai due gruppi: la linea tratteggiata rappresenta la sopravvivenza dei pazienti del Gruppo 1 e la linea intera la sopravvivenza del Gruppo 2. La differenza tra le due curve non è statisticamente significativa, ovvero la presenza del fattore di rischio non ha un impatto statisticamente significativo sulla sopravvivenza.

Test di verifica

1) Chi sono i censurati?

- Sono esclusivamente quei pazienti che arrivano vivi fino alla fine dell'osservazione
- Sono tutti i pazienti che non hanno l'evento di interesse durante il follow-up (ovvero i sopravvissuti, i persi al follow-up, i morti per cause diverse dall'evento di interesse)
- Sono esclusivamente i pazienti persi al follow-up e i sopravvissuti
- Sono esclusivamente i pazienti deceduti per cause diverse dall'evento di interesse.

2) Cosa s'intende per log rank test?

- È un test basato sui logaritmi
- È un test basato sui ranghi
- È un test che serve a confrontare tra loro due curve di sopravvivenza
- È un test che serve a stabilire la forza dell'associazione tra la probabilità del verificarsi di un certo evento ed il tempo di osservazione.

3) In che cosa consiste l'analisi di sopravvivenza con il metodo di Kaplan-Meier?

- a. Il metodo di Kaplan Meier serve a costruire le curve di sopravvivenza di uno o più gruppi di pazienti
- b. L'analisi di sopravvivenza con il metodo di Kaplan-Meier è un particolare tipo di analisi multivariata che tiene conto del tempo
- c. L'analisi di sopravvivenza con il metodo di Kaplan-Meier è un particolare tipo di regressione logistica che permette di mettere in rapporto l'evento con il tempo in cui esso si verifica.

La risposta corretta alle domande sarà disponibile sul sito internet www.sin-italy.org/gin e sul prossimo numero del Giornale cartaceo

Riassunto

L'importanza della statistica nella ricerca biomedica sta nel fatto che essa consente di trarre delle conclusioni o fare inferenze facendo riferimento ad un campione piuttosto che

a tutte le unità costituenti una popolazione. L'analisi di sopravvivenza è parte della statistica inferenziale e la sua peculiarità consiste nel mettere in rapporto un certo esito o evento con il fattore tempo. L'analisi di Kaplan-Meier è il metodo di analisi univariata più largamente utilizzato negli studi di coorte per la costruzione e il confronto delle curve di sopravvivenza.

Indirizzo degli Autori:
Dr. Giovanni Tripepi
CNR-IBIM, Consiglio Nazionale delle Ricerche
Istituto di Biomedicina Epidemiologia Clinica
e Fisiopatologia delle Malattie Renali e
dell'Iperensione Arteriosa – Reggio Calabria
c/o Ki Point – Gransial Srl
Via Filippini, 85
89125 Reggio Calabria
e-mail: gtripepi@ibim.cnr.it