

## I MODELLI ETA' PERIODO COORTE IN EPIDEMIOLOGIA

*Monica Ferraroni*

### 1. Introduzione

L'importanza di distinguere tra gli effetti dell'età (E), del periodo (P) e della coorte di nascita (C) nell'analisi dell'incidenza e della mortalità di una malattia è stata riconosciuta già negli anni '30, quando il problema venne affrontato attraverso un approccio descrittivo che analizzava separatamente i tre effetti.

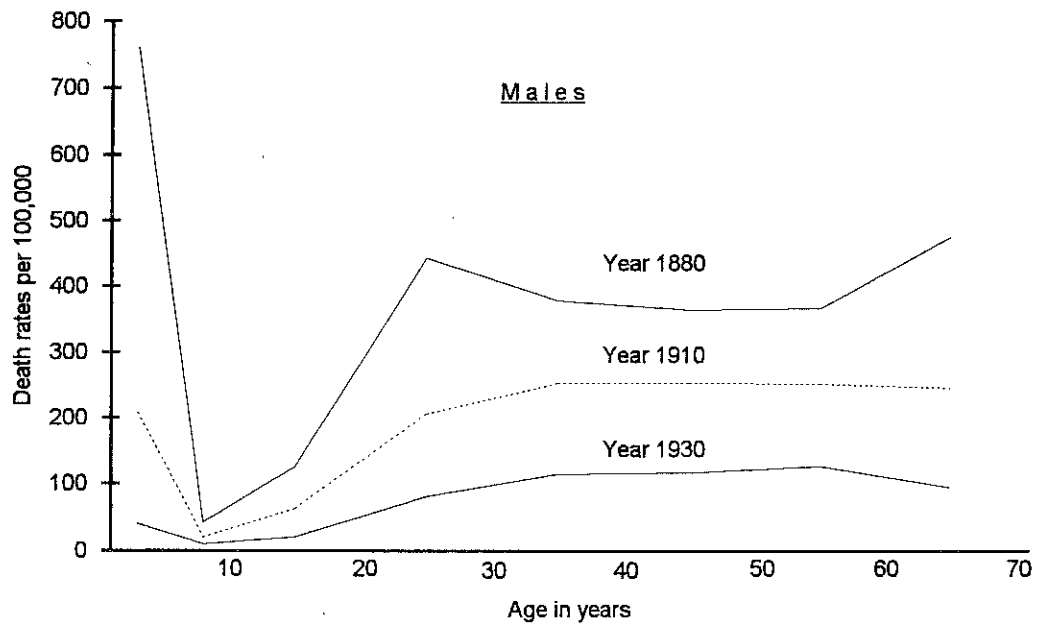
Nel 1939, ad esempio, Forst analizzò l'andamento dei tassi di mortalità specifici per età per tubercolosi nel Massachusetts. Studiando l'andamento dei tassi relativi a tre differenti periodi, il 1880, il 1910 e il 1930, rappresentati in Fig. 1 ripresa dall'articolo originale, egli notava che:

1. per ogni età la mortalità è più bassa nei periodi più recenti;
2. per ogni periodo l'andamento della mortalità è generalmente simile: la mortalità è alta durante l'infanzia, declina nell'età puerile per crescere di nuovo durante l'adolescenza fino a raggiungere livelli più alti durante la vita adulta;
3. nell'ultimo periodo considerato (1930) i tassi di mortalità più alti si trovano tra i 50 e i 60 anni.

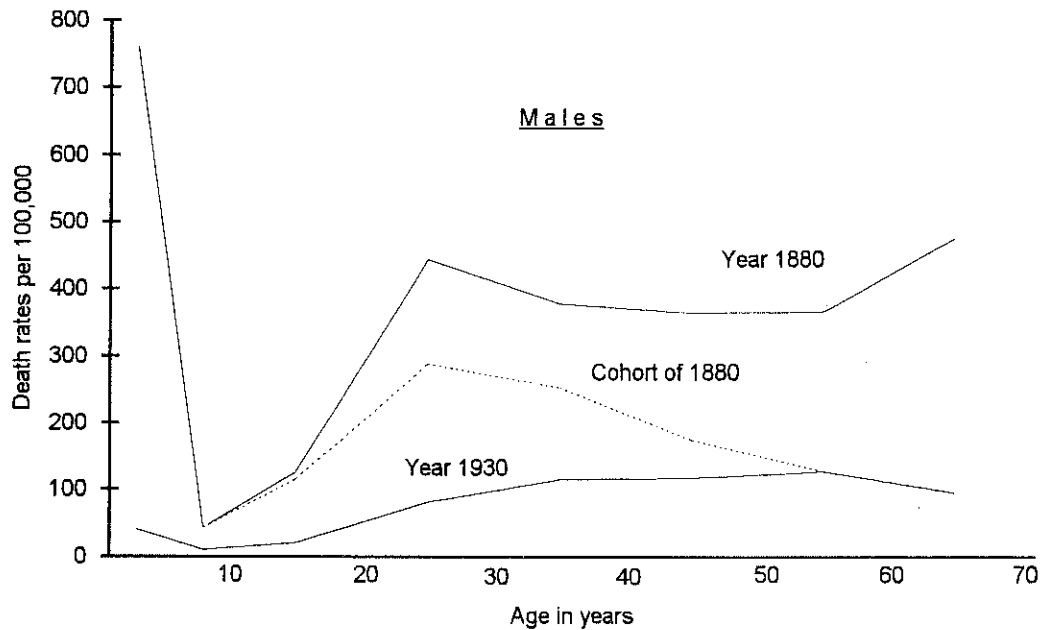
Guardando la curva relativa al 1930, egli scriveva, si ha l'impressione che un individuo incontri il suo più grande rischio di morire di tubercolosi tra i 50 e i 60 anni. In realtà non è realmente così; gli uomini che si trovavano tra i 50 e i 60 anni nel 1930 avevano superato, precedentemente, un maggiore rischio di mortalità.

Questo concetto Forst cercava di metterlo in evidenza attraverso la Fig. 2, che riporta i tassi di mortalità specifici per età per i periodi 1880 e 1930 e, contemporaneamente, l'andamento della coorte del 1880, o meglio del gruppo di persone nate tra il 1871 e il 1880. Si può notare che i soggetti con età tra i 0 e 9 anni nel 1880 hanno, se sopravvissuti, tra i 50 e i 60 anni nel 1930. Questi soggetti prima di affrontare i rischi legati al periodo 1930 sono precedentemente passati attraverso rischi più grandi. Inoltre, la selezione per età nella coorte del 1880 è leggermente diversa rispetto a quella apparentemente indicata dai tassi di mortalità specifici per età per ogni singolo anno.

**Fig. 1: Massachusetts death rates from tuberculosis - all forms - by age, 1880, 1910, 1930 (da Forst, 1939).**



**Fig 2: Massachusetts death rates from tuberculosis - all forms - by age, in the years 1880 and 1930 and for the cohort of 1880 (da Forst, 1939).**



Queste analisi portavano Forst a concludere come fosse altamente probabile che i continui cambiamenti nei tassi di mortalità per la tubercolosi fossero dovuti a cambiamenti

della resistenza umana rispetto al contagio e che i picchi di mortalità nell'età anziana non fossero altro che un residuo di rischi più alti in periodi precedenti.

Negli ultimi anni sono stati proposti modelli statistici usualmente basati su relazioni di tipo lineare, per distinguere gli effetti della coorte di nascita, il periodo di calendario in cui si è verificato l'evento di interesse e l'età del soggetto al momento dell'evento. In questi modelli, l'evento osservato, usualmente rappresentato sotto forma di tassi di mortalità o di incidenza, viene esplicitato in termini di una combinazione lineare di variabili (generalmente età, periodo e coorte o una funzione di esse) e di parametri da stimare di un modello probabilistico. L'approccio basato su modelli ha portato all'unificazione della metodologia per l'analisi dei tassi in epidemiologia descrittiva, per analisi di studi di coorte e per l'analisi di studi di tipo caso-controllo. A fronte di questo indubitabile vantaggio, la modellizzazione probabilistica, non essendo di immediata comprensione, può portare ad interpretazioni non corrette del fenomeno oggetto di studio. Di qui la necessità per gli utilizzatori di questi tipi di modelli, di comprendere appieno le caratteristiche e i limiti di questo tipo di approccio, soprattutto quando utilizzato a scopo inferenziale.

## 2. La rappresentazione dei dati

L'esperienza di mortalità o di incidenza di una popolazione per una determinata causa è normalmente rappresentata attraverso una tabella di contingenza a due vie, in cui le righe rappresentano le classi di età in cui è stata raggruppata la popolazione ( $i=1, \dots, I$ ) e le colonne i periodi di tempo, usualmente di ampiezza pari a quella delle classi di età, in cui la stessa popolazione è stata osservata ( $j=1, \dots, J$ ). Ogni cella può contenere una o più informazioni, quali il numero di eventi accaduti ( $O_{ij}$ ) in soggetti appartenenti alla classe di età  $i$  e nel corso del periodo  $j$  e il numero di anni-persona a rischio ( $N_{ij}$ ), dato dalla somma del numero di anni in cui ogni soggetto, appartenente alla classe  $i$ , è stato osservato nel corso del periodo  $j$  o il rapporto tra queste due quantità ( $r_{ij} = O_{ij} / N_{ij}$ ), che rappresenta il tasso specifico per età cercato. In seguito, per semplicità espositiva, si farà sempre riferimento a situazioni in cui l'evento è rappresentato dal numero di decessi per una definita causa e quindi il tasso è di mortalità. In particolare, quando non altrimenti specificato, si utilizzeranno i dati riportati da Clayton e Schiffers (1987) relativi ai tassi di mortalità specifici per età per 100.000 persone-anno a rischio per tumore della mammella in Giappone nel periodo 1955-1979.

La Tab. 1 rappresenta un tipico esempio di tabella di mortalità. Le coorti di nascita possono essere direttamente ricavate dalla tabella. Ad esempio, la coorte più anziana considerata in questa tabella è quella che identifica i soggetti che muoiono nel periodo 1955-1959 nella classe di età 75-79. Tali soggetti risultano infatti nati nel periodo 1875-1884. La seconda coorte di nascita (1880-1889) è invece stata osservata sia nel periodo 1955-1959, quando apparteneva alla classe di età 70-74, sia nel periodo 1960-1964, quando apparteneva alla classe di età 75-79. Il numero totale di coorti presenti nella tabella è dato da  $K=I+J-1$  mentre  $k=i+j$  rappresenta la coorte di nascita corrispondente all' $i$ -esima classe di età e al  $j$ -esimo periodo di morte, dato che le coorti generalmente

vengono numerate partendo dalla più anziana. Si può notare che le coorti di nascita non solo non sono indipendenti dalle classi di età e dal periodo di morte, in quanto la somma dell'anno di nascita e dell'età alla morte identifica univocamente l'anno di morte, ma, nel caso di dati raggruppati in classi di età, esse si sovrappongono tra loro.

**Tab. 1: Tassi di mortalità specifici per età (per 100.000 anni-persona a rischio) relativi al tumore della mammella in Giappone nel periodo 1955-1979 (Clayton and Schiffers, 1987).**

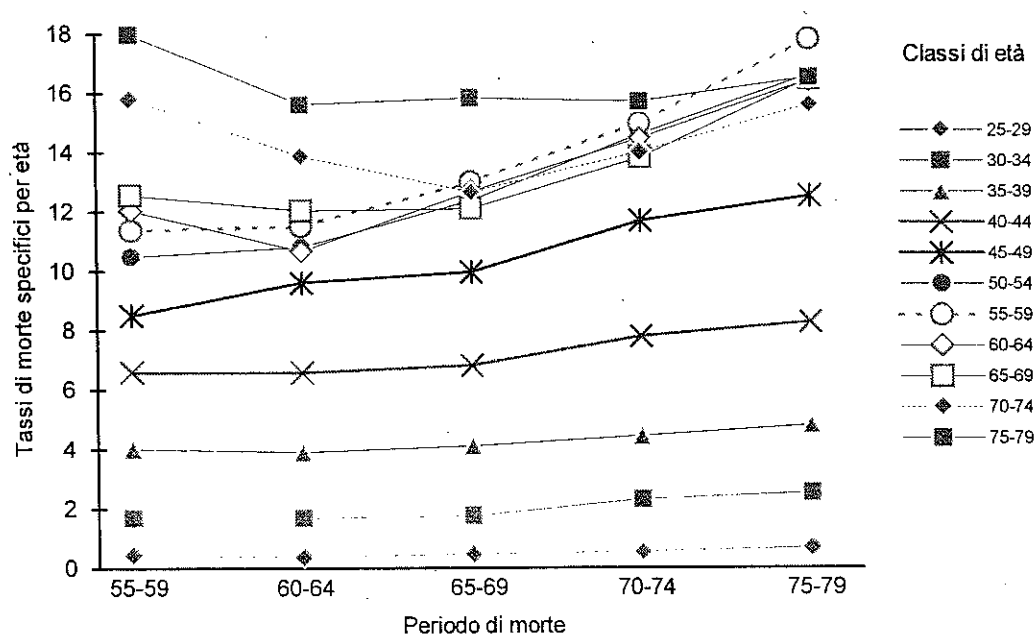
Periodo	1955-1959		1960-1964		1965-1969		1970-1974		1975-1979	
Età										
25-29	0.44	(88)	0.38	(78)	0.46	(101)	0.55	(127)	0.68	(179)
30-34	1.69	(299)	1.69	(330)	1.75	(363)	2.31	(509)	2.52	(588)
35-39	4.01	(596)	3.90	(680)	4.11	(798)	4.44	(923)	4.80	(1056)
40-44	6.59	(874)	6.57	(962)	6.81	(1171)	7.79	(1497)	8.27	(1716)
45-49	8.51	(1022)	9.61	(1247)	9.96	(1429)	11.68	(1987)	12.51	(2398)
50-54	10.49	(1035)	10.80	(1258)	12.36	(1560)	14.59	(2079)	16.56	(2794)
55-59	11.36	(970)	11.51	(1087)	12.98	(1446)	14.97	(1828)	17.79	(2465)
60-64	12.03	(820)	10.67	(861)	12.67	(1126)	14.46	(1549)	16.42	(1962)
65-69	12.55	(678)	12.03	(738)	12.10	(878)	13.81	(1140)	16.46	(1683)
70-74	15.81	(640)	13.87	(628)	12.65	(656)	14.00	(900)	15.60	(1162)
75-79	17.97	(497)	15.62	(463)	15.83	(536)	15.71	(644)	16.52	(865)

La tradizionale rappresentazione grafica di questi dati è quella in cui i tassi di morte specifici per età vengono rappresentati in ordinata come funzione del periodo di morte del soggetto o dell'età al decesso, secondo l'interesse del ricercatore. In Fig. 3 i tassi vengono presentati in funzione del periodo di morte, i punti rappresentano quindi, per le varie classi di età alla morte, l'andamento nei vari periodi. La peculiarità più evidente di questa rappresentazione è il generale parallelismo degli andamenti dei tassi di morte tra i periodi per ogni età, con valori via via crescenti per le età più anziane.

L'andamento dei tassi di morte può essere anche analizzato per ogni specifica generazione o coorte di nascita. Gli effetti di coorte in questo caso vengono rappresentati rileggendo la Tab. 1 in modo longitudinale, considerando cioè ogni sua diagonale come una colonna di una nuova tabella in cui per riga vengono rappresentate le classi di età alla morte e per colonna le coorti di nascita. Questa nuova rappresentazione viene riportata in Tab. 2 e in Fig. 4 in cui l'ascissa rappresenta il valore centrale della classe di età alla morte e i punti congiunti, come indicato in legenda, rappresentano alcune delle 15 coorti di nascita presenti in Tab. 2.

In scala logaritmica l'andamento delle curve riportate in Fig. 4 risulta parallelo. Questo sembra indicare che i tassi di mortalità specifici per età di ogni coppia di coorti di nascita stanno fra loro approssimativamente in un rapporto costante per l'intero arco della vita.

Fig. 3: Tassi di mortalità specifici per età (per 100.000 anni-persona a rischio) relativi al tumore della mammella in Giappone nel periodo 1955-1979. Effetto Età-Periodo.



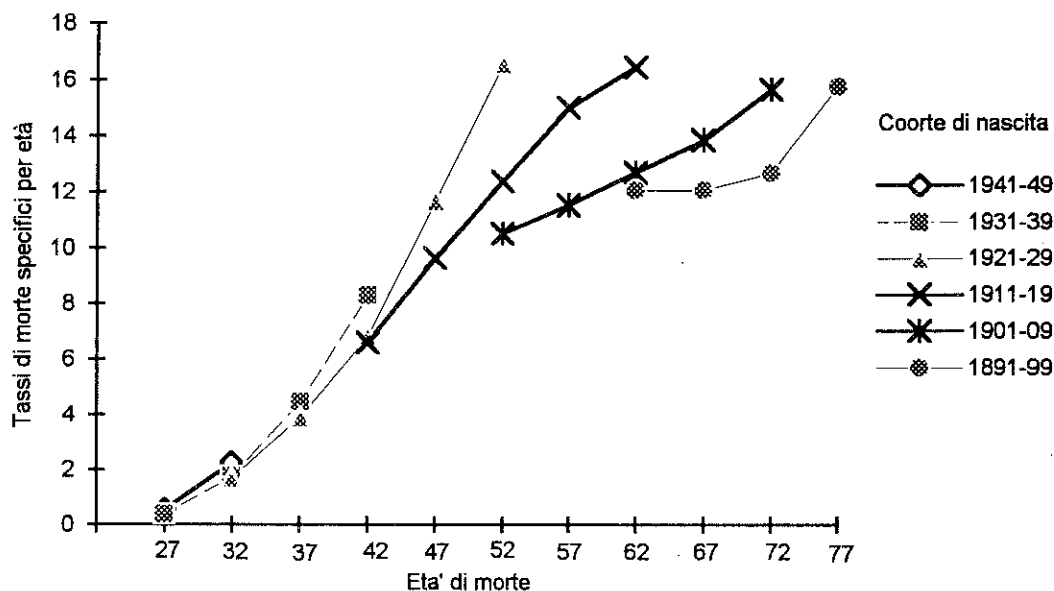
Tab. 2: Tassi di mortalità specifici per età relativi al tumore della mammella in Giappone nel periodo 1955-1979. Rilettura della Tab. 1 in senso longitudinale per evidenziare le coorti di nascita di cui si riporta solo l'anno centrale.

Coorte	1880	1885	1890	1895	1900	1905	1910	1915	1920	1925	1930	1935	1940	1945	1950
Età															
25-29											0.44	0.38	0.46	0.55	0.68
30-34										1.69	1.69	1.75	2.31	2.25	
35-39									4.01	3.90	4.11	4.44	4.80		
40-44								6.59	6.57	6.81	7.79	8.27			
45-49						8.51	9.61	9.96	11.68	12.51					
50-54					10.49	10.80	12.36	14.59	16.56						
55-59				11.36	11.51	12.98	14.97	17.79							
60-64			12.03	10.67	12.67	14.46	16.42								
65-69		12.55	12.03	12.10	13.81	16.46									
70-74	15.81	13.87	12.65	14.00	15.60										
75-79	17.97	15.62	15.83	15.71	16.52										

Entrambe le rappresentazioni esplicitate graficamente attraverso le Fig. 3 e 4 possono trovare una descrizione funzionale attraverso modelli probabilistici in grado, nel primo caso, di separare l'effetto dell'età dall'effetto del periodo (modello età-periodo) e nel secondo l'effetto dell'età dall'effetto della coorte (modello età-coorte). Questi modelli in generale ipotizzano che una qualsiasi funzione del tasso atteso possa essere espressa, su

una scala lineare, come somma di contributi delle singole componenti, età, periodo e coorte.

**Fig. 4: Tassi di mortalità specifici per età (per 100.000 anni-persona a rischio) relativi al tumore delle mammella in Giappone nel periodo 1955-1979. Effetto Età-Coorte.**



In particolare, se  $y$  è una funzione del tasso specifico per età atteso, generalmente il logaritmo, il modello età-periodo per la  $i$ -esima classe di età e il  $j$ -esimo periodo di morte, può essere matematicamente espresso dalla seguente espressione:

$$y_{ij} = \ln[E(r_{ij})] = \mu + \alpha_i + \pi_j$$

dove  $\mu$  rappresenta l'intercetta,  $\alpha_i$  (con  $i=1,2,\dots,I$ ) l'effetto dell' $i$ -esima classe di età e  $\pi_j$  (con  $j=1,2,\dots,J$ ) rappresenta l'effetto del  $j$ -esimo periodo di calendario. Similmente, il modello età-coorte per la  $i$ -esima classe di età e la  $k$ -esima coorte può essere espresso tramite la seguente:

$$y_{ik} = \ln[E(r_{ik})] = \mu + \alpha_i + \gamma_k$$

dove  $\mu$  rappresenta ancora l'intercetta,  $\alpha_i$  (con  $i=1,2,\dots,I$ ) l' $i$ -esimo effetto dell'età e  $\gamma_k$  (con  $k=1,2,\dots,K$ ) rappresenta il  $k$ -esimo effetto della coorte di nascita.

Nell'ipotesi che il numero di decessi attesi specifici per età e per periodo seguano una distribuzione di tipo poissoniano, la stima degli effetti avviene usualmente mediante il metodo dei minimi quadrati sul logaritmo dei tassi, con pesi uguali al numero di morti osservati, che approssimano l'inverso della varianza del logaritmo del tasso. Alternativamente, può essere utilizzato il metodo della massima verosimiglianza, con ipotesi di distribuzione poissoniana dei decessi osservati o di distribuzione binomiale dei

tassi, anche alla luce delle note proprietà del metodo. La bontà di adattamento del modello viene usualmente valutata verificando che i valori attesi dal modello interpolato non siano troppo diversi da quelli osservati. A questo proposito, come sottolineato da Clayton e Schifflers (1987), occorre notare che la naturale variabilità esistente tra i numeratori dei vari tassi, può comportare che l'insieme dei parametri stimati possa interpolare meglio alcune celle e peggio altre. Risulta allora opportuno valutare attentamente i tassi attesi nelle celle con un maggior numero di eventi in quanto essi influenzano maggiormente qualsiasi processo di stima.

Nei modelli lineari sopra definiti le stime dei parametri non risultano univocamente identificate in quanto esse sono funzione dei vincoli matematici imposti ai parametri stessi. Ciò comporta che lo stesso valore dei tassi attesi può essere ottenuto da più insiemi di parametri. Ad esempio, nel modello età-periodo è sufficiente aggiungere un valore costante ad ogni valore stimato  $\hat{\alpha}_i$  e sottrarre lo stesso valore ad ogni  $\hat{\pi}_j$  stimato perché i valori attesi  $\hat{y}_{ij}$  rimangano invariati. In ogni caso, qualsiasi sia l'insieme di parametri stimati, considerati due periodi,  $j$  e  $j+1$ , per ogni classe di età, la differenza tra il logaritmo dei tassi è data da:

$$y_{ij+1} - y_{ij} = \pi_{j+1} - \pi_j$$

e può essere interpretata come il logaritmo del rischio relativo del periodo  $j+1$  rispetto a  $j$ . Similmente, date le due classi di età  $i+1$  e  $i$ , la differenza  $(\alpha_{i+1} - \alpha_i)$  può essere interpretata come il rischio relativo di decesso nei soggetti appartenenti alla classe di età  $i+1$  rispetto a quelli appartenenti alla classe  $i$ . Di conseguenza, benché non sia identificabile l'insieme di parametri, risulta identificabile la differenza fra essi in scala logaritmica, o il loro rapporto nella scala originale.

In campo epidemiologico, la parametrizzazione di questi modelli viene usualmente scelta in modo tale che le stime degli effetti rappresentino quantità interpretabili. Ad esempio, fissare uno degli effetti periodo ( $\pi_j$ ) o degli effetti coorte ( $\gamma_k$ ) a zero, equivale ad esprimere i risultati, in un modello in scala logaritmica, in termini di differenze rispetto al  $j$ -esimo periodo (o alla  $k$ -esima coorte). In questo modo gli  $\alpha_i$  stimano direttamente il logaritmo dei tassi di mortalità specifici per età per il periodo (o la coorte) usato come riferimento mentre i restanti parametri  $\pi$  (o  $\gamma$ ) rappresentano il logaritmo del rischio relativo per quel periodo (o quella coorte) rispetto al periodo (o coorte) di riferimento. Altri tipi di parametrizzazioni possono essere utilizzate per risolvere il problema della non identificabilità, ad esempio si possono centrare gli effetti del periodo o della coorte in modo da imporre che la loro media valga zero. A causa della maggiore difficoltà interpretativa, questi tipi di parametrizzazione vengono raramente utilizzati.

In ogni caso, i parametri stimati possono essere attentamente valutati solo dopo aver verificato che il modello ben si adatti ai tassi osservati, e cioè quando, attraverso un attento esame dei residui, non si sia evidenziata la presenza di celle male interpolate o di insiemi di valori che possono indicare qualche tipo di errore sistematico o di violazione degli assunti distribuzionali del modello.

In generale il modello che meglio interpreta dati di mortalità o di incidenza è il modello età-coorte. Relativamente ai dati riportati in Tab. 1 ad esempio, il modello età-periodo presenta una devianza uguale a 203 con 40 gradi di libertà (gdl) mentre il modello

età-coorte ha una devianza di 81 con 30 gdl. Anche se nessuno dei due modelli è significativo, il modello età-coorte sembra comunque meglio interpolare i dati come evidenziato dal valore inferiore della devianza. Questo notevole miglioramento in devianza è comunque ottenuto tramite un modello più complesso (10 parametri in più, pari alla differenza tra i gdl dei due modelli) in quanto in Tab. 1 sono considerati 5 periodi e 15 coorti. C'è da sottolineare inoltre che la stima degli effetti di alcune coorti risulta meno affidabile in quanto basata su meno eventi di altre. Tipicamente è il caso delle ultime coorti reclutate che vengono osservate solo per gli intervalli di età più giovani e delle più anziane che vengono osservate solo nel primo periodo. Confrontare questi due modelli significa quindi confrontare modelli con complessità diversa e di conseguenza, se da un lato non sorprende che il modello età-coorte interpreti meglio i dati, dall'altro occorre interpretarne con più cautela i risultati.

A volte può comunque succedere che un insieme di tassi venga ugualmente ben descritto dai due modelli, e cioè che contemporaneamente esistano un effetto periodo ed un effetto coorte significativi.

**Tab. 3: Tassi di mortalità specifici per età (per 100.000 anni-persona a rischio) relativi al tumore del polmone nelle donne Belga durante il periodo 1955-1978. Il numero di morti è riportato in parentesi (Clayton and Schiffers, 1987).**

Periodo	1955-1959	1960-1964	1965-1969	1970-1974	1975-1978
Età					
25-29	0.19 (3)	0.13 (2)	0.50 (7)	0.19 (3)	0.70 (10)
30-34	0.66 (11)	0.98 (16)	0.72 (11)	0.71 (10)	0.57 (7)
35-39	0.78 (11)	1.32 (22)	1.47 (24)	1.64 (25)	1.32 (15)
40-44	2.67 (36)	3.16 (44)	2.53 (42)	3.38 (53)	3.93 (48)
45-49	4.84 (77)	5.60 (74)	4.93 (68)	6.05 (99)	6.83 (88)
50-54	6.60 (106)	8.50 (131)	7.65 (99)	10.59 (142)	10.42 (134)
55-59	10.36 (157)	12.00 (184)	12.68 (189)	14.34 (180)	17.95 (177)
60-64	14.76 (193)	16.37 (232)	18.00 (262)	17.60 (249)	23.91 (239)
65-69	20.53 (219)	22.60 (267)	24.90 (323)	24.33 (325)	32.70 (343)
70-74	26.24 (223)	27.70 (250)	30.47 (308)	36.94 (412)	38.47 (358)
75-79	33.47 (198)	33.61 (214)	36.77 (253)	43.69 (338)	45.20 (312)

In questo caso gli effetti del periodo e quelli della coorte generalmente evidenziano un trend lineare nei logaritmi di uguale entità; ad esempio, nei tassi specifici per età relativi alla mortalità per tumore al polmone delle donne Belga durante i periodi 1955-1978 riportati in Tab. 3, si evidenzia, come indicato in Tab. 4, una mortalità crescente quasi monotona ad un tasso medio di circa il 10% per ogni periodo o coorte considerati.

E' forse utile ricordare che non è possibile confrontare direttamente il modello età-periodo ed il modello età-coorte costruendo un test formale per verificare quale dei due modelli sia significativamente migliore dell'altro.



**Tab. 4: Mortalità per tumore al polmone tra le donne Belga. Modello età-periodo ed età-coorte.**

Età+Periodo (Devianza=38.2, gdl=40)				Età+Coorte (Devianza=29.5, gdl=30)			
Età	$\alpha_i$	Periodo	$\pi_j$	Età	$\alpha_i$	Coorte	$\gamma_k$
25-29	-12.82	1955-1959	0.000	25-29	-13.54	1880	-0.331
30-34	-12.01	1960-1964	0.107	30-34	-12.42	1885	-0.318
35-39	-11.43	1965-1969	0.162	35-39	-11.85	1890	-0.231
40-44	-10.58	1970-1974	0.278	40-44	-10.92	1895	-0.105
45-49	-9.99	1975-1978	0.423	45-49	-10.22	1900	0.000
50-54	-9.55			50-54	-9.66	1905	0.055
55-59	-9.12			55-59	-9.14	1910	0.203
60-64	-8.83			60-64	-8.73	1915	0.331
65-69	-8.50			65-69	-8.29	1920	0.470
70-74	-8.25			70-74	-7.94	1925	0.484
75-79	-8.07			75-79	-7.67	1930	0.655
						1935	0.740
						1940	0.717
						1945	0.361
						1950	1.664

### 3. Il modello Età-Periodo-Coorte

Qualora né il modello età-periodo né quello età-coorte riescano a spiegare la variabilità nei tassi osservati, allora è possibile ipotizzare un modello più complesso comprendente simultaneamente i tre effetti:

$$y_{ij} = \mu + \alpha_i + \pi_j + \gamma_k + \varepsilon_{ij} \quad (1)$$

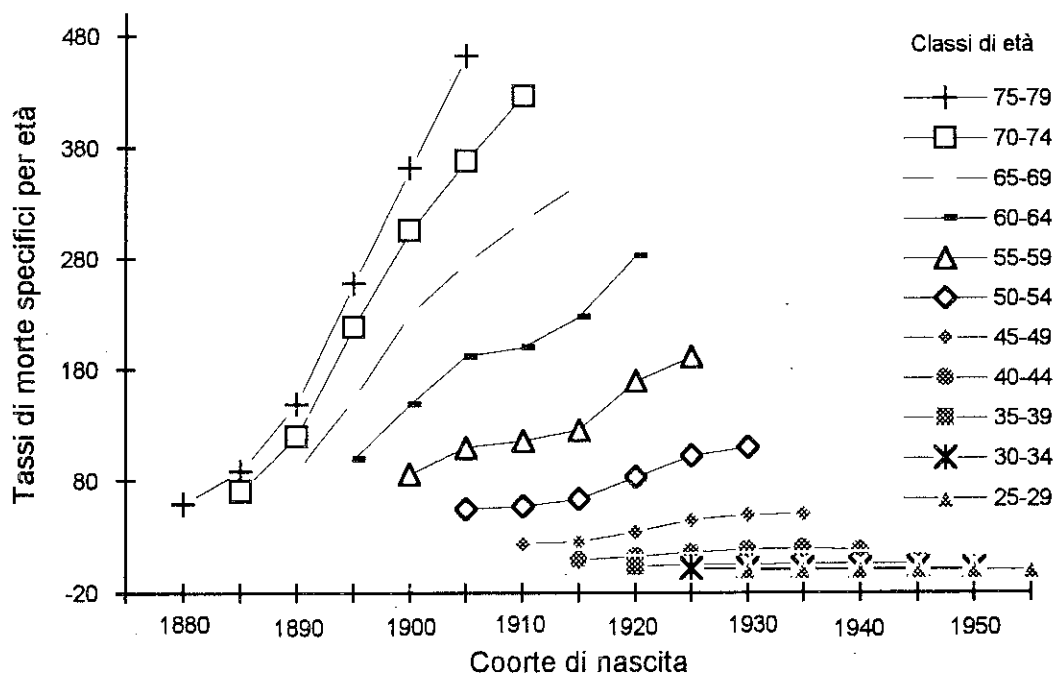
in cui  $\mu$  rappresenta l'intercetta,  $\alpha_i$ ,  $\pi_j$  e  $\gamma_k$  rappresentano, come in precedenza, gli effetti dell'età, del periodo e della coorte rispettivamente e la variabile dipendente  $y_{ij}$  è una funzione del tasso oggetto di studio,  $y_{ij} = f[E(O_{ij} / N_{ij})]$ .

In questo contesto i modelli età-periodo e età-coorte si configurano come casi particolari di modelli afferenti alla classe più generale dei modelli età-periodo-coorte.

Negli anni più recenti, modelli appartenenti a questa classe sono stati utilizzati come momento intermedio per il confronto dell'esperienza di mortalità/morbilità di coorti diverse. In quest'ambito vanno collocate, ad esempio, procedure quali quella utilizzata all'International Agency for Research on Cancer (Coleman et al., 1993) basata sul confronto di tassi cumulativi per le diverse coorti, valutati partendo dai valori attesi dei tassi in accordo ad un definito modello. Queste procedure di uso sempre più frequente in epidemiologia descrittiva, risultano in ogni caso più facilmente interpretabili quando siano chiari i limiti sottesi ai modelli della classe età-periodo-coorte utilizzati per la stima dei tassi cumulativi. Per questo in questa nota l'attenzione sarà posta propriamente sui modelli età-periodo coorte piuttosto che sull'uso che delle stime da essi derivanti possa essere fatto.

In campo epidemiologico le ragioni che portano a privilegiare modelli in grado di valutare gli effetti congiunti delle tre componenti sono eminentemente euristici. In tale settore infatti le varie componenti possono avere distinte interpretazioni biologiche legate al fenomeno oggetto di studio. Per molti tumori nell'uomo, ad esempio, intercorre sovente un intervallo di tempo, generalmente alcuni decenni, tra la prima esposizione ad un cancerogeno, e lo sviluppo clinico della malattia o la morte a causa di essa. Questo comporta che la mortalità per tumore odierna è fortemente influenzata da abitudini od esposizioni che i soggetti avevano contratto diversi anni prima. Tipico esempio è il tumore al polmone dove la mortalità per questa patologia può venire in larga misura spiegata in termini di variazione nel consumo di sigarette. In Fig. 5 sono riportati gli andamenti dei tassi di mortalità specifici per età ( $\times 100.000$ ) per tumore alla trachea, bronchi e polmoni (ICD 162) in Italia tra il 1955 e il 1984 relativamente ad ogni anno di nascita (La Vecchia et al., 1989).

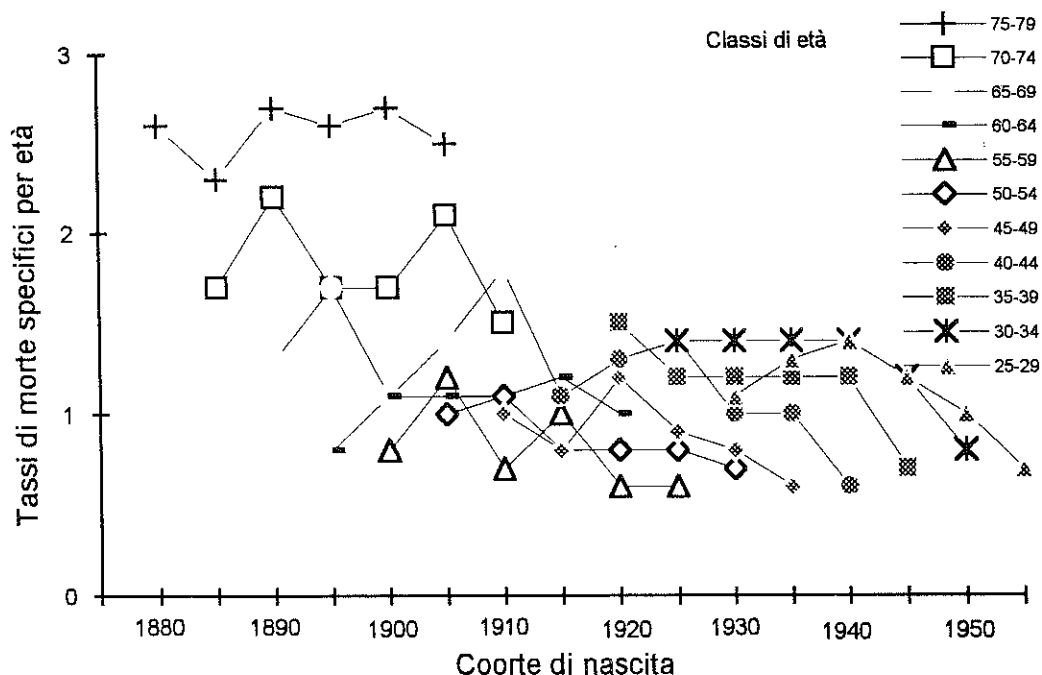
**Fig. 5: Andamento dei tassi di mortalità specifici per età (per 100.000 anni persona a rischio) relativi al tumore alla trachea, bronchi e polmoni (ICD 162) in Italia nel periodo 1955-1984.**



Vi sono comunque dei fattori che possono influenzare la mortalità per tumori nel breve periodo indipendentemente dalla coorte di nascita. Ad esempio, l'introduzione di una nuova terapia efficace per una determinata neoplasia tenderebbe a far diminuire, nel breve periodo, la mortalità per quello stesso tumore in tutte le classi di età. Uno degli esempi più noti è il tumore al testicolo, dove l'introduzione di nuove terapie chemioterapiche all'inizio degli anni '70 ha quasi annullato la mortalità per questa

neoplasia, sempre che venga diagnosticata in tempo. In Fig. 6 sono riportati gli andamenti dei tassi di mortalità specifici per età ( $\times 100.000$ ) per tumore al testicolo (ICD 186) in Italia tra il 1955 e il 1984 relativamente ad ogni anno di nascita (La Vecchia et al., 1989). Analogamente, vi sono dei cancerogeni che agiscono su uno degli ultimi stadi del processo di carcinogenesi la cui introduzione può velocemente provocare un aumento nell'incidenza o mortalità per tumori. Quindi, se l'identificazione di un effetto coorte potrebbe indicare l'esposizione a fattori in grado di agire su uno degli stadi iniziali del processo di carcinogenesi oppure l'acquisizione di abitudini diversificate legate a modifiche generali di tipo socioeconomico, un effetto periodo potrebbe indicare sia l'esposizione a carcinogeni con azione che si sviluppa sugli stadi finali della malattia (promuovente del tumore) sia eventuali miglioramenti sul piano diagnostico e/o terapeutico.

**Fig. 6: Andamento dei tassi di mortalità specifici per età (per 100.000 anni persona a rischio) relativi al tumore del testicolo (ICD 186) in Italia nel periodo 1955-1984.**



Pertanto, un'analisi che consenta di distinguere l'effetto coorte di nascita dall'effetto periodo di morte risulta in campo epidemiologico, ed in particolare nell'analisi dell'incidenza e della mortalità, di notevole interesse sia dal punto di vista dello studio dei meccanismi dell'eziologia della malattia, sia per valutare l'impatto che innovazioni diagnostiche e terapeutiche possono avere sull'incidenza e sulla mortalità.

Il principale problema derivante dalla modellizzazione dei tre tipi di effetto è legato alla relazione di tipo lineare esistente tra età, periodo e coorte, come risulta evidente dalla (1). Infatti, la somma dell'anno di nascita e dell'età in cui è avvenuto il decesso identifica univocamente l'anno di morte.

La dipendenza lineare tra i fattori determina infatti un problema di non identificabilità da un punto di vista matematico, in quanto ogni insieme di parametri è legato funzionalmente ad un altro tramite una trasformazione lineare del tipo:

$$\alpha_i + \pi_j + \gamma_k \equiv (\alpha_i + c \cdot i) + (\pi_j - c \cdot j) + (\gamma_k - c \cdot (i - j))$$

dove  $c$  è una costante arbitraria. Questo significa che l'inclusione di una qualsiasi componente lineare in uno dei fattori porta a modificare i parametri stimati per gli altri fattori, senza peraltro cambiare la stima dei tassi attesi.

Diverse soluzioni sono state proposte per superare questo tipo di problema, anche in considerazione del fatto che normalmente si è maggiormente interessati al confronto dell'esperienza tra due o più gruppi (coorti, periodi) che all'esatta quantificazione dei vari effetti, cioè al valore assoluto della stima dei parametri.

James e Segal (1982) ad esempio, hanno proposto un modello non lineare del tipo:

$$\log[E(\epsilon_{ij})] = \mu + \pi_j + \gamma_k + \alpha_i \cdot \delta_j + \epsilon_{ij}$$

dove  $\pi_j$  e  $\gamma_k$  rappresentano ancora gli effetti periodo e coorte, mentre l'effetto età  $\alpha_i$  viene incluso congiuntamente ad un fattore moltiplicativo  $\delta_j$  funzione del periodo. Modelli di questo tipo, quando basati su solide ipotesi biologiche relative alla causa in studio, risultano di estremo interesse, in quanto permettono una stima univoca dei parametri. D'altra parte però, i modelli non lineari comportano l'uso di metodi di stima più complessi dal punto di vista computazionale e producono usualmente stime poco stabili.

L'approccio proposto da Holford (1983), ed in seguito ripreso da Capocaccia (1986), si colloca anch'esso nell'ambito della modellizzazione statistica dei tassi con distribuzione del numero di eventi di tipo poissoniano, ed in particolare nella classe delle funzioni stimabili dei parametri che sono invarianti rispetto ad un particolare vincolo. Nel suo lavoro Holford propone di scindere ogni effetto dovuto all'età, al periodo ed alla coorte in due componenti. La prima che definisce un trend lineare e la seconda che quantifica la curvatura locale del trend cioè di quanto l'effetto devia dalla linearità. Supponendo che il vincolo scelto per il modello (1) sia  $\sum_i \alpha_i = \sum_j \pi_j = \sum_k \gamma_k = 0$ , e quindi che i rispettivi

effetti età, periodo e coorte siano centrati rispetto al loro valore medio, il trend lineare per, ad esempio, l'effetto età può essere dato da:

$$\alpha_L = C \sum_i c_i \alpha_i$$

dove  $c_i = i - 1/2$  e  $C = \left( \sum_i c_i^2 \right)^{-1}$ , mentre la componente di curvatura è data dall'effetto età quando il trend lineare è rimosso, e cioè  $\tilde{\alpha}_i = \alpha_i - c_i \alpha_L$ .

In modo simile vengono scomposti gli effetti del periodo e quello della coorte. Per parametrizzare tutti questi effetti viene così creata una matrice del disegno che contiene i coefficienti necessari a stimare i sei effetti previsti, data da  $\mathbf{X} = (\mathbf{1} \mathbf{A}_c \mathbf{P}_c \mathbf{C}_c \mathbf{A}_L \mathbf{P}_L \mathbf{C}_L)$ . Per come è costruita, la matrice non risulta di rango pieno in quanto  $\mathbf{C}_L = \mathbf{P}_L - \mathbf{A}_L$ . Il processo di stima è quindi ottenibile tramite un'inversa generalizzata, che non ha una univoca soluzione. La funzione lineare  $\mathbf{q}'\beta$ , dove  $\beta' = (\mu \alpha'_c \pi'_c \gamma'_c \alpha_L \pi_L \gamma_L)$  rappresenta

il vettore di coefficienti relativi alla matrice del disegno, è stimabile se  $q'\beta = t \log \lambda$  per ogni  $t$ , dove  $\log \lambda$  rappresenta il logaritmo dei tassi. (Searle, 1971). La proprietà che qui si rivela di particolare interesse è che la curvatura locale dei trend risulta una funzione stimabile invariante rispetto alla particolare parametrizzazione usata. Di conseguenza, essa può essere interpretata in modo univoco, a differenza di quanto avviene per le componenti lineari che risultano interpretabili solo con le cautele sopra menzionate. In altri termini, mentre non risulta possibile quantificare in modo univoco un eventuale aumento costante nei tassi all'aumentare del periodo, in quanto tale aumento può essere dovuto o ad un effetto coorte o al fatto che la coorte in studio semplicemente invecchia, risulta comunque possibile stimare una eventuale curvatura locale cioè qualsiasi scostamento dall'andamento lineare dei tassi.

Il metodo recentemente proposto da Tarone e Chu (1996), comprende come caso particolare il metodo proposto da Holford e si inserisce nel capitolo della interpretazione degli effetti età, periodo e coorte attraverso l'uso di funzioni stimabili. Esso parte dal presupposto che benché non siano matematicamente identificabili le pendenze dei vari trend lineari, risultano comunque identificabili le differenze in pendenza; possono cioè essere definiti particolari contrasti tra le varie coorti (o periodi) che, risultando invarianti rispetto alla parametrizzazione scelta, portano alla stima di parametri identificabili. In questo approccio non parametrico è sempre possibile definire contrasti in grado di stimare la curvatura locale dei trend, cioè contrasti che portano alle stesse stime ottenibili con il metodo di Holford.

L'uso di vincoli sui parametri che possono essere sia arbitrari che basati su conoscenza a priori rispetto al fenomeno oggetto di studio, è il sistema più utilizzato in epidemiologia per effettuare l'analisi di dati di mortalità o di incidenza in quanto l'interpretazione dei parametri risulta di particolare semplicità. Tra i vari modelli di questo tipo presenti in letteratura, quello proposto da Osmond e Gardner (1982) ed in seguito ripreso da Decarli e La Vecchia (1987) è forse quello che ha trovato maggiori applicazioni ed è quello che verrà qui presentato in dettaglio.

Supponendo che il numero di eventi (decessi o diagnosi) avvenuti nella  $i$ -esima classe di età e nel  $j$ -esimo periodo di osservazione ( $O_{ij}$ ) abbiano distribuzione di tipo poissoniano e che  $y_{ij} = E(O_{ij} / N_{ij})$ , l'equazione (1) può essere riscritta nel seguente modo:

$$\log[E(O_{ij})] = \log N_{ij} + \log a_i + \log p_j + \log c_k$$

dove  $\log a_i$ ,  $\log p_j$  e  $\log c_k$  assumono lo stesso significato di  $\alpha_i$ ,  $\pi_j$  e  $\gamma_k$  visto in precedenza.

La stima dei parametri può essere effettuata con il metodo dei minimi quadrati pesati, cioè minimizzando la seguente funzione:

$$f(a, p, c) = \sum_{i,j} O_{ij} (\log y_{ij} - \log a_i - \log p_j - \log c_k)^2 \quad (2)$$

dove i pesi sono rappresentati da  $O_{ij}$ , cioè da valori pari all'inverso della varianza campionaria dei logaritmi dei tassi. Il minimo di  $f(a,p,c)$  è trovato eguagliando a zero le sue derivate prime rispetto ai parametri oggetto di stima. Le derivate parziali di  $f(a,p,c)$

rispetto a  $\log a_i$ ,  $\log p_j$  e  $\log c_k$  producono un insieme di I+J+K equazioni normali in I+J+K variabili. Infatti:

$$\frac{\partial f(a, p, c)}{\partial \log a_i} = \frac{\partial f(a, p, c)}{\partial \log p_j} = \frac{\partial f(a, p, c)}{\partial \log c_k} = -2 \sum_{i,j} O_{ij} (\log y_{ij} - \log a_i - \log p_j - \log c_k)$$

Ponendo = 0 queste I+J+K equazioni si trova il minimo della funzione f, e cioè:

$$\sum_{i,j} O_{ij} (\log y_{ij} - \log a_i) = \sum_j \log p_j \left( \sum_i O_{ij} \right) + \sum_k \log c_k \left( \sum_{(i,j): k=i+j} O_{ij} \right)$$

Questa relazione suggerisce tre possibili vincoli, quali:

$$\begin{aligned} \sum_{i,j} (\log y_{ij} - \log a_i) O_{ij} &= 0 \\ \sum_j \log p_j \left( \sum_i O_{ij} \right) &= 0 \\ \sum_k \log c_k \left( \sum_{(i,j): k=i+j} O_{ij} \right) &= 0 \end{aligned}$$

che hanno una soluzione banale per  $p_0 = 1_j$ ,  $c_0 = 1_k$  e  $a_0 = (\sum_j O_{ij} \log y_{ij}) / \sum_j O_{ij}$ .

I vincoli posti sui parametri portano, da un punto di vista geometrico, alla definizione di un modello che tende a descrivere l'andamento dei tassi tramite una curva ben definita e un insieme di spostamenti rigidi applicati alla curva stessa. In particolare, il vincolo posto per  $a_0$  porta alla stima di una curva che risulta essere l'andamento medio della mortalità per età nei vari periodi considerati, mentre le stime degli effetti periodo/coorte quantificano l'entità delle modifiche da apportare a questa curva media per considerare la sua evoluzione nel tempo. Questo tipo di vincoli privilegia quindi una interpretazione basata sugli effetti periodo e coorte. Simmetricamente possono essere definiti modelli che privilegiano interpretazioni età/periodo o età/coorte definendo vincoli diversi su  $c_0$  e  $p_0$  rispettivamente.

Dalla relazione lineare esistente tra gli effetti dell'età, del periodo e della coorte, consegue che se  $\log a_i$ ,  $\log p_j$  e  $\log c_k$  sono un insieme di parametri in grado di minimizzare la (2), anche:

$$\begin{aligned} \log a'_i &= \log a_i + \lambda(I - i) \\ \log p'_j &= \log p_j + \lambda \cdot j \\ \log c'_k &= \log c_k - \lambda \cdot k \end{aligned} \quad (3)$$

costituisce una soluzione al problema.

La differenza tra i due insiemi di parametri,  $\log a_i$ ,  $\log p_j$ ,  $\log c_k$  e  $\log a'_i$ ,  $\log p'_j$ ,  $\log c'_k$ , consiste, sul piano geometrico, semplicemente in una rotazione degli effetti, la cui

ampiezza dipende dal valore di  $\lambda$ . Questo significa che la soluzione del modello non è unica; infatti, qualsiasi sia la soluzione ottenuta dal modello basta ruotare l'effetto della coorte in un verso e l'effetto del periodo nel verso opposto per ottenere un'altra soluzione valida. Il modello a tre componenti (età, periodo, coorte) ha quindi un'infinità di soluzioni equivalenti, cioè con stesso valore della devianza residua, una per ogni valore di  $\lambda$ . Poiché ad ogni valore di  $\lambda$  corrispondono interpretazioni diverse dei parametri, non risulta possibile, interpolando tutti e tre gli effetti contemporaneamente, distinguere l'effetto della coorte dall'effetto del periodo. Affinché un parametro sia identificabile, esso deve assumere lo stesso valore per qualunque valore di  $\lambda$ .

Nell'approccio proposto da Tarone e Chu, come già detto, si suppone che la pendenza del trend negli effetti nella coorte di nascita differisca nei diversi periodi in modo tale che:

$$\begin{aligned} \gamma_k &= \theta_1 + \beta_1 k && \text{se } k \leq k_1 \\ \gamma_k &= \theta_2 + \beta_2 k && \text{se } k \geq k_2 \end{aligned}$$

con  $k_1 < k_2$ , dove  $\theta_1$  e  $\theta_2$  sono le intercette dei trend rispettivamente nei due periodi considerati.

La differenza tra i due coefficienti angolari,  $\beta_1 - \beta_2$ , risulta avere lo stesso valore per tutti gli infiniti insiemi di parametri che specificano i valori medi associati ad un determinato modello. Infatti, data la trasformazione:

$$\gamma_k^* = \gamma_k - \lambda k$$

se  $k \leq k_1$  allora  $\gamma_k^* = \theta_1 + \beta_1 k - \lambda k = \theta_1 + (\beta_1 - \lambda)k = \theta_1 + \beta_1^* k$ , e se  $k \geq k_2$  allora  $\gamma_k^* = \theta_2 + \beta_2 k - \lambda k = \theta_2 + (\beta_2 - \lambda)k = \theta_2 + \beta_2^* k$ . Quindi, mentre i coefficienti angolari degli effetti trasformati risultano ancora una volta diversi al variare di  $\lambda$ , la differenza tra i coefficienti angolari risulta uguale e quindi identificabile, infatti  $\beta_2^* - \beta_1^* = \beta_2 - \beta_1$ . In questo modo è possibile effettuare un'inferenza univoca rispetto ai cambiamenti in grandezza o direzione dei trend nelle coorti di nascita. Analogamente è possibile dimostrare che risulta univoca la stima della differenza tra i coefficienti angolari del trend dell'effetto periodo.

Per ottenere una stima della differenza tra i coefficienti angolari nei differenti periodi, occorre definire dei contrasti anch'essi invarianti rispetto alla trasformazione precedentemente definita. In Fig. 7 si presentano le stime di massima verosimiglianza degli effetti di coorte relativi ai tassi standardizzati di mortalità per tumore alla mammella negli USA tra il 1970 e il 1989 riportati nell'articolo di Tarone e Chu in cui le stime di massima verosimiglianza sono ottenute imponendo la prima e l'ultima coorte di nascita come punto di riferimento.

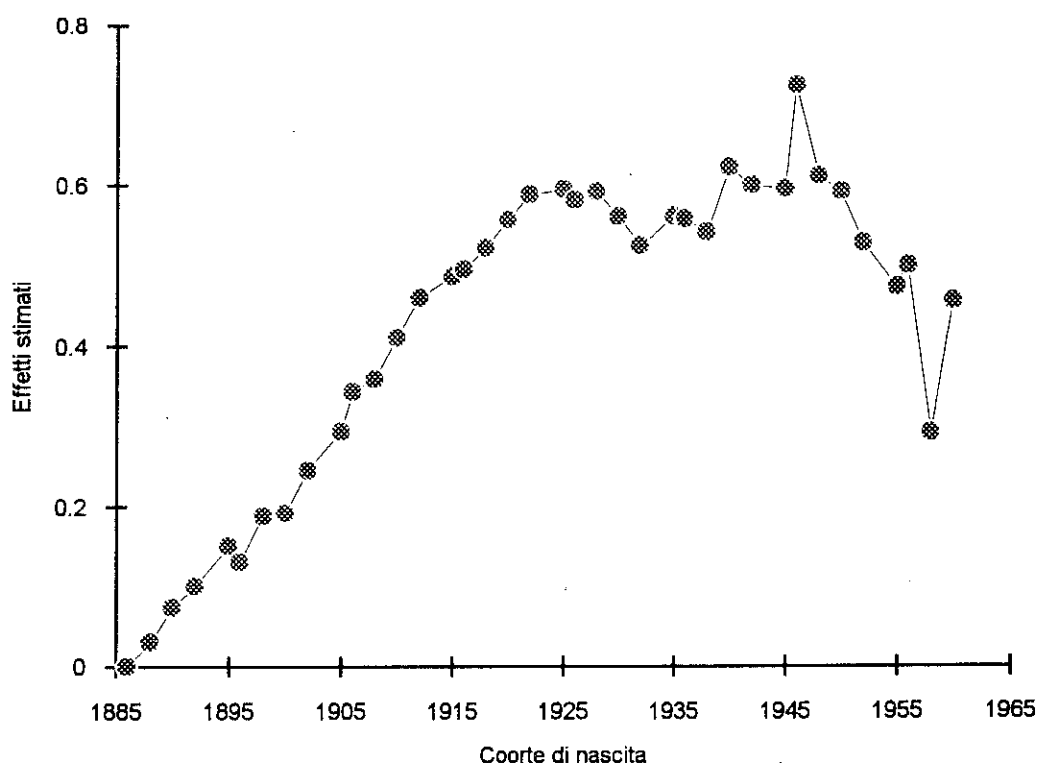
Partendo dall'ipotesi che singoli parametri hanno sempre un'interpretazione relativa, gli autori contrastano due gruppi di coorti (o di periodi) dove sembra esserci in modo evidente un trend differente. Ad esempio, da analisi precedenti risultava un trend crescente significativo nelle coorti di nascita dal 1900 al 1914 ed un trend significativamente decrescente dal 1924 al 1938. Per confrontare le pendenze dei trend lineari si usano i seguenti contrasti:

$$C_1 = \gamma_{k+7} - \gamma_k - (\gamma_{h+7} - \gamma_h)$$

$$C_2 = 7\gamma_{k+7} + 5\gamma_{k+6} + 3\gamma_{k+5} + \gamma_{k+4} - \gamma_{k+3} - 3\gamma_{k+2} - 5\gamma_{k+1} - 7\gamma_k - (7\gamma_{h+7} + 5\gamma_{h+6} + 3\gamma_{h+5} + \gamma_{h+4} - \gamma_{h+3} - 3\gamma_{h+2} - 5\gamma_{h+1} - 7\gamma_h)$$

dove  $h+7 \leq k$ .  $C_1$  consente di verificare la stima della differenza delle due pendenze considerando solo la coorte iniziale e la coorte finale nei due gruppi, mentre  $C_2$  verifica la stima della differenza delle due pendenze considerando ogni singola coorte nel gruppo. Ognuno di questi contrasti può essere espresso algebricamente come una combinazione lineare dei parametri  $C = s'\gamma$  dove  $s$  è il vettore dei coefficienti che definiscono il contrasto e  $\gamma$  è il vettore degli effetti di coorte.

**Fig. 7: Stima degli effetti coorte del modello età-periodo-coorte per il tumore alla mammella nelle donne USA, 1970-1989. Modello di Tarone e Chu.**



Volendo contrastare i due gruppi di coorti sopra indicati occorre definire  $h=8$ , che corrisponde alla coorte 1901 e  $k=20$ , corrispondente alla coorte 1925. Il contrasto  $C_1$  vale -0.353 con un errore standard di 0.018; mentre il contrasto  $C_2$  vale -4.157 con errore standard stimato di 0.175. Entrambi i contrasti indicano un cambiamento di pendenza altamente significativo.



Il problema di identificazione del modello, e cioè della stima univoca dei parametri, è stato risolto da Osmond e Gardner (1982) con una procedura che si è rivelata di particolare interesse nello studio della mortalità e incidenza di malattie.

Dapprima si interpolano tre modelli log-lineari a due fattori, il modello età-periodo, il modello età-coorte e il modello periodo-coorte, cioè si minimizzano le tre funzioni che si possono ricavare dalla (2) utilizzando in modo appropriato i vincoli precedentemente descritti. Interpolare il modello età-periodo ad esempio, è l'equivalente di minimizzare l'equazione (2) rispetto l'età e il periodo ponendo come vincolo che  $c_0 = 1_k$  mentre interpolare il modello periodo-coorte significa minimizzare la (2) rispetto al periodo e alla coorte imponendo che  $a_0 = (\sum_j O_{ij} \log y_{ij}) / \sum_j O_{ij}$ .

Ogni modello fornisce un vettore contenente le stime dei parametri ( $X$ ) e una devianza residua che indica quanto il modello ben si adatta ai dati, e cioè:

$X_a = (a_0, \hat{p}_a, \hat{c}_a)$	$G_a^2$	dal modello periodo-coorte
$X_p = (\hat{a}_p, p_0, \hat{c}_p)$	$G_p^2$	dal modello età-coorte
$X_c = (\hat{a}_c, \hat{p}_c, c_0)$	$G_c^2$	dal modello età-periodo
$X(\lambda) = (\hat{a}, \hat{p}, \hat{c}, \lambda)$	$G^2$	dal modello età-periodo-coorte

Il modello completo, per quanto detto finora, fornisce a sua volta un vettore di stime che sono parametrizzabili rispetto all'ignoto parametro  $\lambda$ , qui indicate con  $X(\lambda)$ . La differenza tra le soluzioni fornite dai modelli a due fattori e il modello completo può essere misurata dalla distanza euclidea valutata nello spazio  $\mathcal{R}^{1+J+K}$  esistente tra i vettori stimati. Indicando con  $\| \cdot \|$  la distanza euclidea, per ogni modello a due fattori si ha:

$$d_a(\lambda) = \|X_a - X(\lambda)\|$$

$$d_p(\lambda) = \|X_p - X(\lambda)\|$$

$$d_c(\lambda) = \|X_c - X(\lambda)\|$$

Il metodo proposto da Osmond e Gardner, tra tutte le possibili soluzioni  $X(\lambda)$  sceglie quella corrispondente al valore  $\lambda^*$  tale per cui la somma delle distanze  $d_a(\lambda)$ ,  $d_p(\lambda)$  e  $d_c(\lambda)$  risulti minima. In questa procedura di minimizzazione ognuno dei tre modelli a due fattori ha peso inversamente proporzionale al valore della statistica  $G^2$  che misura la bontà di adattamento del corrispondente modello divisa per i rispettivi gradi di libertà, e cioè:

$$g(\lambda) = \frac{d_a(\lambda)}{(G_a^2 / (I-2) \cdot (J-1))} + \frac{d_p(\lambda)}{(G_p^2 / (I-1) \cdot (J-2))} + \frac{d_c(\lambda)}{(G_c^2 / (I-1) \cdot (J-1))}$$

Minimizzando la funzione  $g(\lambda)$  si ottiene quindi un valore  $\lambda^*$  che definisce il vettore finale delle stime degli effetti del modello età-periodo-coorte  $X(\lambda^*)$ . Le stime  $X(\lambda^*)$  così ottenute saranno maggiormente influenzate dai modelli a due fattori che

meglio interpolano i dati. Inoltre, la minimizzazione di  $g(\lambda)$  crea un meccanismo di compromesso rispetto alla parametrizzazione in  $\lambda$  delle stime del modello età-periodo-coorte dove, per quanto visto nella (3), l'effetto periodo cresce all'aumentare di  $\lambda$  mentre diminuisce l'effetto coorte. Scegliere il valore che minimizza  $g(\lambda)$  significa non dare nessuna speciale enfasi a nessun particolare effetto. Ad esempio, se i tassi sono crescenti (o decrescenti) attraverso l'intera tabella, questo si rifletterà sia nei valori del periodo sia in quelli della coorte mentre piccoli cambiamenti nella matrice dei tassi rifletterà piccoli cambiamenti nel valore dei parametri. La confidenza nei valori stimati dei parametri dipende da quanto grande è la differenza che esiste tra i valori di  $\lambda$  che minimizzano  $d_a(\lambda)$ ,  $d_p(\lambda)$  e  $d_c(\lambda)$  rispettivamente e il valore trovato  $\lambda^*$ .

Inoltre è stata resa disponibile una procedura GLIM in cui, pur utilizzando i vincoli per il modello saturo come accennato precedentemente, le stime degli effetti vengono normalizzate rispetto ad un effetto medio generale; mentre le stime degli effetti dell'età rimangono i tassi specifici per età, la stima degli effetti del periodo e della coorte sono invece dei rischi relativi rispetto ad un periodo o ad una coorte media.

I dati relativi al tumore alla mammella delle donne Giapponesi riportati in Tab. 1 portano ad un  $G^2$  per il modello età-periodo-coorte di 30.7 con 27 gdl. Le stime degli effetti sono riportate in Tab. 5.

**Tab. 5: Mortalità per tumore alla mammella nelle donne Giapponesi nel periodo 1955-1979. Stima degli effetti del modello Età+Periodo+Coorte ottenute con macro GLIM (Decarli e La Vecchia).**

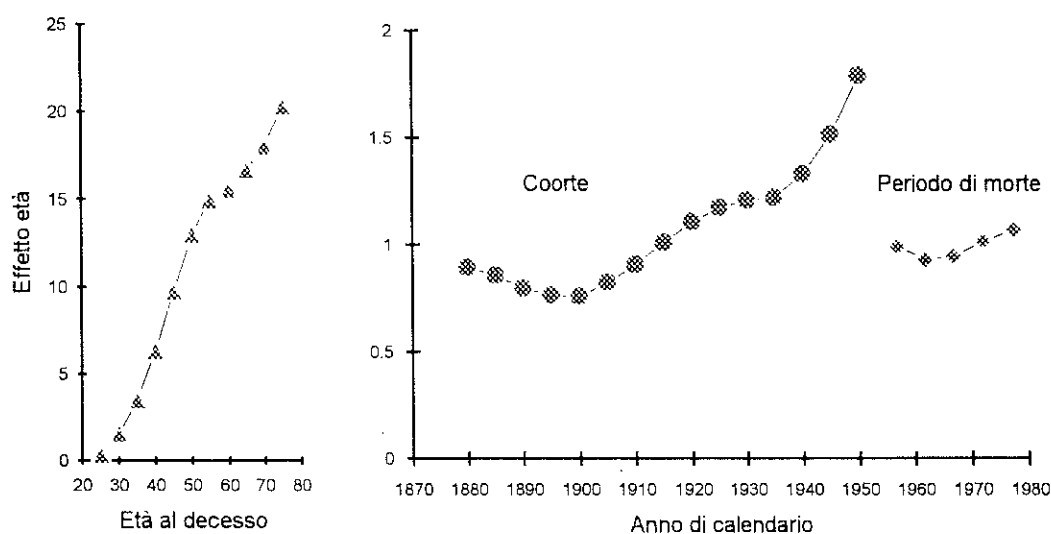
Età+Periodo+Coorte (Devianza=30.7, gdl=27)					
Età	$\alpha_i$	Periodo	$\pi_j$	Coorte	$\gamma_k$
25-29	0.36	1955-1959	0.989	1880	0.895
30-34	1.56	1960-1964	0.931	1885	0.859
35-39	3.54	1965-1969	0.944	1890	0.799
40-44	6.36	1970-1974	1.016	1895	0.765
45-49	9.76	1975-1978	1.065	1900	0.759
50-54	13.02			1905	0.825
55-59	14.98			1910	0.908
60-64	15.60			1915	1.009
65-69	16.70			1920	1.107
70-74	18.08			1925	1.173
75-79	20.32			1930	1.206
				1935	1.218
				1940	1.327
				1945	1.514
				1950	1.786

Come si può facilmente notare nessun effetto viene usato come riferimento. Il rischio di sviluppare l'evento nel periodo 1975-1978 rispetto al primo periodo disponibile (1955-1959) può essere facilmente valutato facendo il rapporto tra i due parametri stimati

$1.065/0.989=1.0768$ . Volendo invece valutare se l'essere nati in questo secolo o nel precedente può rappresentare un rischio, la coorte del 1900 risulta un utile riferimento. Ad esempio, appartenere alla coorte del 1950 rispetto a quella del 1900 comporta un rischio relativo di  $1.786/0.759=2.3531$ .

Gli andamenti delle stime sono invece riportate in Fig. 8. L'andamento dell'età indica un aumento dei tassi specifici per età, all'aumentare dell'età a cui si verifica l'evento, dovuto al fatto che i tassi rappresentano la mortalità per un tumore tipico delle popolazioni anziane. Sembra comunque esserci una lieve attenuazione della pendenza della curva per età attorno alla menopausa tra i 55 e i 60 anni.

**Fig. 8: Stime degli effetti del modello età-periodo-coorte per il tumore alla mammella nelle donne Giapponesi nel periodo 1955-1979.**

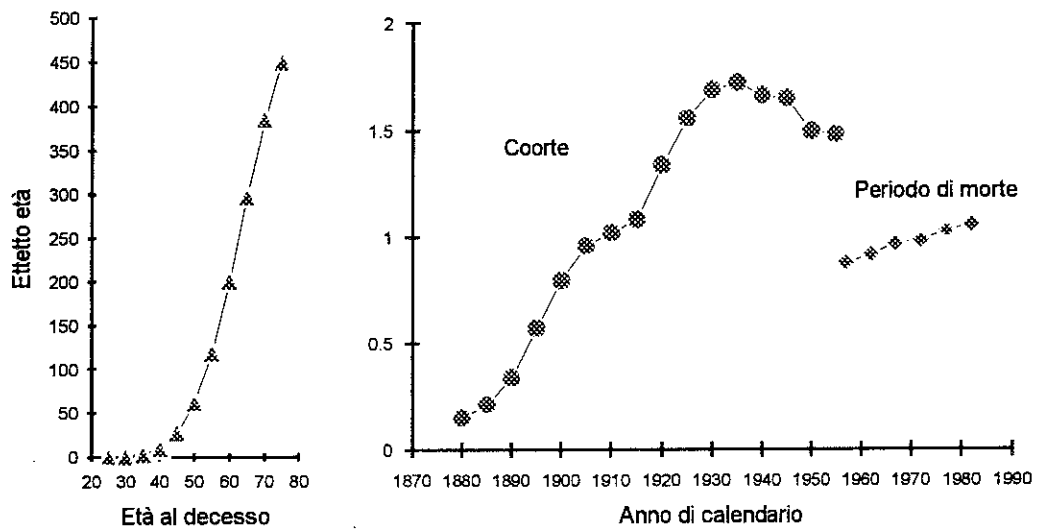


L'andamento del periodo rivela un declino del rischio fino agli anni '60 ed una successiva ripresa. Un andamento ad U dell'effetto periodo potrebbe indicare o un cambiamento del modo di codificare le cause di morte o un effettivo aumento all'esposizione a fattori di rischio di questa malattia. Come effetto coorte si può notare un andamento abbastanza costante fino alla quartultima coorte, corrispondente ai nati verso il 1935, ed una successiva brusca ripresa della mortalità per tumore alla mammella.

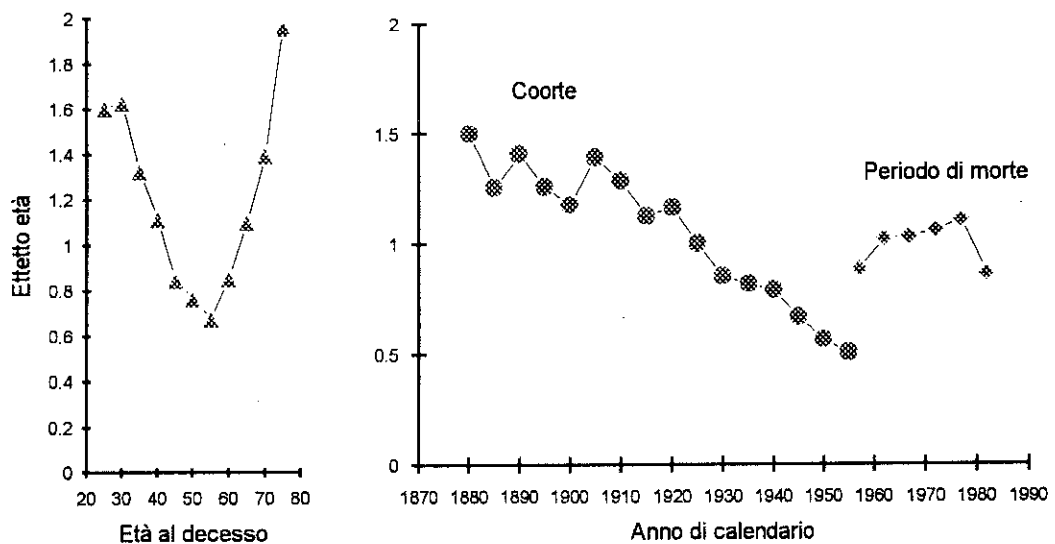
Per quanto riguarda l'andamento delle stime degli effetti del modello età-periodo-coorte per il tumore al polmone e il tumore ai testicoli nei maschi italiani dal 1955 al 1984, esse sono riportate in Fig. 9 e 10 rispettivamente.

Per il tumore al polmone la mortalità è aumentata notevolmente nelle generazioni nate tra il 1880 e il 1905 e, successivamente, tra il 1915 ed il 1930 mentre i valori relativi al periodo di morte sono aumentati costantemente. Gli andamenti osservati possono venir spiegati dalla variazione del consumo di sigarette, che ha mostrato due grossi incrementi dopo le due guerre mondiali.

**Fig. 9: Stime degli effetti del modello età-periodo-coorte per il tumore al polmone nei maschi italiani nel periodo 1955-1984.**



**Fig. 10: Andamenti delle stime degli effetti del modello età-periodo-coorte per il tumore al testicolo nei maschi italiani nel periodo 1955-1984.**



Il fatto che si siano osservati aumenti continui sia nei valori riferiti alla coorte che al periodo di morte può venir considerato un'ulteriore conferma della duplice azione del fumo di sigarette sulla carcinogenesi bronchiale, poiché l'influenza su uno dei primi stadi implicherebbe un effetto coorte e l'azione su uno degli ultimi un effetto periodo. I maschi italiani nati dopo il 1940, d'altra parte, mostrano una diminuzione del loro rischio di

morte per tumore al polmone, che riflette le recenti diminuzioni nella prevalenza del fumo di sigarette negli uomini.

La curva per età del tumore al testicolo presenta un picco nella terza decade di vita ed una successiva diminuzione fino a 50 anni, per poi risalire nell'età più anziana. Questo riflette la presenza di diversi tipo istologici per questo tipo di tumore. Le stime degli effetti coorte e periodo riflettono invece l'interazione tra i due fattori, poiché soltanto le generazioni più recenti hanno potuto godere degli avanzamenti terapeutici registrati a partire dagli anni '70 che emergono chiaramente dalla recente diminuzione delle stime per periodo.

#### 4. Riferimenti bibliografici

Capocaccia R., "Variazioni temporali della mortalità in Italia" in *Atti 3° Convegno Nazionali sugli Studi di Mortalità*, Lega Italiana per la Lotta contro i Tumori, Firenze 1986.

Clayton D., Schifflers E., "Models for temporal variation in cancer rates. I. Age-period and age-cohort models" *Stat Med*, 6, 1987, pp. 449-467.

Clayton D., Schifflers E., "Models for temporal variation in cancer rates. II. Age-period-cohort models" *Stat Med*, 6, 1987, pp. 468-481.

Coleman MP., Estève J., Damiecki P., Arslan A., Renard H., *Trends in cancer incidence and mortality*, IARC Scientific Publications No 121, Lyon, 1993.

Decarli A., La Vecchia C., "Age, period and cohort models: review of knowledge and implementation in GLIM" *Rivista di Statistica Applicata*, 20, 1987, pp. 397-410.

Frost WH., "The age selection of mortality from tuberculosis in successive decades" *Am J of Hygiene*, 30A, 1939, pp. 91-96.

Holford TR., "The estimation of age, period and cohort effects for vital rates" *Biometrics*, 39, 1983, pp. 311-324.

James IR, Segal MR., "On a method of mortality analysis incorporating age-year interaction, with application to prostate cancer mortality" *Biometrics*, 38, 1982, pp. 311-324.

La Vecchia C, Negri E, Decarli A, Fasoli M, Cislighi C., "Cancer mortality in Italy: an overview of age-specific and age standardised trends from 1955 to 1984" *Tumori*, 76, 1989, pp. 87-166.

Osmond C., Gardner MJ., "Age, period and cohort models applied to cancer mortality rates" *Stat Med*, 1, 1982, pp. 245-259.

Searle SR., *Linear Models*, Wiley, New York, 1971.

Tarone RE., Chu KC., "Evaluation of birth cohort patterns in population disease rates" *Am J Epidemiol*, 143, 1996, pp. 85-91.

