

## Metodi statistici e definizioni

Si illustrano di seguito alcuni degli indicatori utilizzati nelle statistiche presenti sui rapporti presentati, le modalità di calcolo e i riferimenti bibliografici.

### *Classi di età.*

In accordo con le popolazioni mostrate nelle Tabelle 1, i dati sono stati accorpati per convenienza in 18 classi di età. Per esempio, denominando con  $i$  la singola classe di età, allora quando  $i$  vale 1 si farà riferimento alla prima classe di età (ovvero tra 0 e 4 anni) mentre quando  $i$  vale 18, si farà riferimento alla classe dei soggetti con 85 e più anni.

Questa divisione in 18 classi di età è di fatto uno standard condiviso dalla maggior parte dei registri tumori internazionali e consente di tenere in considerazione il fattore età senza sostanziali distorsioni.

### *Numero casi.*

Indica il numero di nuove diagnosi di tumore (incidenza) o di decessi per tumore (mortalità) nel periodo considerato.

Dato  $n_i$  il numero di nuovi casi/decessi nella fascia di età generica  $i$ , si avrà che il numero totale dei nuovi casi o di decessi,  $N$ , sarà dato da:

$$N = \sum_{i=1}^{18} n_i$$

### *Popolazione.*

Analogamente, data  $p_i$  la numerosità della popolazione nella fascia di età generica  $i$ , si avrà che la popolazione totale  $P$  sarà:

$$P = \sum_{i=1}^{18} p_i$$

### *Tasso età specifico.*

Il tasso età specifico  $t_i$  indica il numero di casi per abitante; nelle statistiche riportate si usa una costante moltiplicativa pari a 100.000 (qui omesso), in questo modo si ottiene il numero di casi ogni 100.000 abitanti.

$$t_i = \frac{n_i}{p_i}$$

la costante 100.000 è un valore standard che facilita la leggibilità del valore del tasso di malattia ed è usata comunemente nell'ambito dell'epidemiologia dei tumori negli adulti. Per i tumori infantili, molto più rari, si usa solitamente la costante moltiplicativa 1.000.000 (casi su milione).

### *Tasso grezzo.*

Il numero di nuovi casi per abitante è denominato tasso grezzo  $TG$  ed è anch'esso espresso per 100.000.

$$TG = \frac{\sum_{i=1}^{18} n_i}{\sum_{i=1}^{18} p_i} = \frac{N}{P}$$

vogliamo tuttavia sensibilizzare il lettore nell'usare questo indicatore con estrema cautela in quanto, pur tenendo in considerazione le dinamiche della popolazione nel suo complesso, non è sensibile alla diversa distribuzione delle fasce di età che, nell'epidemiologia tumorale, sono determinanti.

#### *Tasso standardizzato diretto.*

Esprime una misura di sintesi dei tassi età specifici che, specie se raffrontata con il tasso grezzo, consente un più agevole confronto tra aree diverse. Di fatto i tassi standardizzati sono la misura di incidenza e mortalità di riferimento per confronti nel tempo (andamenti temporali) e nello spazio (confronti tra aree) in quanto contengono un aggiustamento per età.

Alla base del calcolo del tasso standardizzato con cosiddetto metodo diretto (*TSD*) vi è la definizione di una serie di pesi  $w_i$  per ogni  $i$ -esima classe di età.

$$TSD = \frac{\sum_{i=1}^{18} w_i t_i}{\sum_{i=1}^{18} w_i}$$

Nella Tabella 2 si mostrano alcuni pesi comunemente usati per il calcolo dei tassi standardizzati diretti (Dos Santos Silva, 1999 e Curado, 2005).

Età	Europa	Mondo
0-4	0,08	0,12
5-9	0,07	0,1
10-14	0,07	0,09
15-19	0,07	0,09
20-24	0,07	0,08
25-29	0,07	0,08
30-34	0,07	0,06
35-39	0,07	0,06
40-44	0,07	0,06
45-49	0,07	0,06
50-54	0,07	0,05
55-59	0,06	0,04
60-64	0,05	0,04
65-69	0,04	0,03
70-74	0,03	0,02
75-79	0,01	0,01
80-84	0,01	0,005
85+	0,01	0,005
<b>Totale</b>	<b>1</b>	<b>1</b>

Tabella 2. *Pesi standard per il tasso standardizzato.*

Nelle statistiche presentate si fa uso della sola standardizzazione secondo la popolazione standard europea, indicata nella prima colonna della Tabella, in quanto rappresenta lo standard *de facto* per la presentazione dei dati di incidenza tumorale in Italia soprattutto ma anche in Europa.

La standardizzazione secondo la popolazione standard mondiale è tuttavia comunemente usata, specie nei confronti internazionali, qui omessi. Come sempre tutti i tassi sono espressi per 100.000 abitanti.

### *Tasso standardizzato indiretto.*

Un metodo di standardizzazione che non utilizza pesi esterni (come invece accade nella standardizzazione diretta) è quello della standardizzazione indiretta.

Vengono in questo caso applicati dei tassi attesi età specifici  $e_i$ :

$$SIR/SMR/SPR = \frac{N}{\sum_{i=1}^{18} e_i p_i}$$

Il rapporto standardizzato di incidenza (SIR) o mortalità (SMR) o di prevalenza (SPR) è quindi espresso come rapporto tra il numero di casi osservati e il numero di casi attesi, ovvero se i tassi della popolazione in esame fossero quelli riscontrati nella popolazione di confronto.

Quando il SIR/SMR/SPR supera l'unità (o il valore 100, quando, come nei rapporti presentati, questa misura viene espressa in percentuale) significa che nella popolazione osservata vi è stato un eccesso di casi rispetto ai tassi di confronto. Un difetto invece se la misura del SIR/SMR/SPR fosse inferiore a 1 (o 100 nel caso di numero percentuale).

Ovviamente con piccole numerosità questi eccessi o difetti possono essere puramente casuali o non significativi. Associamo pertanto l'indicazione, grafica o numerica, dei limiti di confidenza del SIR/SMR/SPR. Se entrambi i limiti fossero in aree di eccesso o di difetto, allora si potrebbe concludere che tale eccesso o difetto sia da considerarsi statisticamente significativo (Breslow, 1987).

### *Prevalenza.*

La prevalenza è la misura dei soggetti con diagnosi di tumore e viventi ad una data di riferimento. La prevalenza è un indicatore di grande e specifica utilità per la pianificazione sanitaria in quanto indicativa del reale carico assistenziale che necessitano i pazienti oncologici.

La misura riportata riguarda la prevalenza dei pazienti diagnosticati negli ultimi 10 anni, ovvero in altre parole: i pazienti viventi ad una certa data e che hanno avuto una diagnosi di tumore fino a 10 anni prima quella stessa data. In alcuni casi è possibile assumere che i pazienti viventi ma con diagnosi avvenuta in periodi precedenti i 10 anni siano guariti e quindi non più pazienti oncologici.

Una corretta stima della prevalenza limitata al semplice conteggio dei pazienti viventi ad una certa data potrebbe indurre una certa sottostima a causa della proporzione, generalmente esigua, dei soggetti per i quali non è noto lo stato in vita.

Si applica pertanto una formula per la stima dei casi prevalenti: in particolare il numero di casi prevalenti alla fine dell'anno  $t$  con diagnosi effettuata negli  $n$  anni precedenti,  $P_n(t)$ , viene stimato nel seguente modo:

$$P_k(t) = \sum_{j=t-n+1}^t \left[ l_j(t) \left( 1 + \frac{w_j(t)}{l_j(t) + d_j(t)} \right) \right]$$

Dove  $l_j(t)$  è il numero di soggetti incidenti diagnosticati nell'anno  $j$  e viventi alla fine dell'anno  $t$ ,  $w_j(t)$  è il numero di soggetti incidenti diagnosticati nell'anno  $j$  e senza informazioni sullo stato in vita alla fine dell'anno  $t$  e  $d_j(t)$  è il numero di soggetti incidenti diagnosticati nell'anno  $j$  e deceduti prima della fine dell'anno  $t$ .

Esistono formulazioni alternative che implicano il calcolo della sopravvivenza attesa (AIRTUM Working Group, 2010) queste formulazioni producono risultati sovrapponibili a quelli da noi ottenuti in questi rapporti.

### Sopravvivenza.

La sopravvivenza esprime una misura dei pazienti che sopravvivono alla malattia a distanza di tempo (in genere anni) dalla diagnosi.

Una misura della sopravvivenza elementare è la sopravvivenza osservata  $SO$ :

$$SO_k = 1 - \prod_{j=1}^k \frac{d_j}{l_j - w_j / 2}$$

con  $SO_k$  la sopravvivenza osservata dopo  $k$  anni dalla diagnosi,  $d_j$  il numero di decessi nell'anno  $j$ ,  $l_j$  il numero di soggetti viventi all'inizio dell'anno  $j$ ,  $w_j$  il numero di soggetti persi al follow-up nel corso dell'anno  $j$ .

Di fatto la sopravvivenza osservata misura la proporzione di pazienti viventi a distanza di tempo dalla diagnosi. Questa semplice formulazione ha lo svantaggio di tenere in considerazione tutti i decessi, anche quelli non correlati al tumore in oggetto. In questo modo si ottiene una sovrastima della letalità del tumore in quanto si attribuisce implicitamente ad esso l'intera mortalità dei pazienti. Questa sovrastima è tanto maggiore quanto maggiore è l'età del paziente.

Nell'ambito dei registri tumori la misura più utilizzata è quindi un'altra, la sopravvivenza relativa  $SR$ , calcolata come:

$$SR_k = \frac{SO_k}{SA_k}$$

Dove  $SA_k$  è la sopravvivenza attesa della popolazione generale con le medesime caratteristiche (età e genere) dei pazienti considerati. La sopravvivenza relativa è calcolata secondo un metodo chiamato Ederer II (AIRTUM Working Group, 2011), ed è basata sulle tavole di mortalità per età e sesso dell'area in esame.

### Criteri classificativi

La definizione delle sedi presenti nei rapporti presentati è basata sulla terza revisione della Classificazione Internazionale delle Malattie per l'Oncologia (ICD-O-3) e alla nona edizione della Classificazione Internazionale delle Malattie (ICD-9). Si rimanda a questi volumi per maggiori dettagli (Fritz, 2000 e Jones, 1994).

il dettaglio dei criteri utilizzati per raggruppare le neoplasie nei rapporti presentati sulla base di queste due classificazioni sono riportati nella tabella riepilogativa (Tabella 3).

Si noti principalmente la separazione del rene dalle altre vie urinarie, e la suddivisione dell'encefalo/sistema nervoso centrale e della vescica in considerazione del comportamento (maligno e non maligno).

Tabella 3. Criteri per la definizione delle neoplasie

DESCRIZIONE	TOPOGRAFIA ICD-O-	MORFOLOGIA ICD-O-3	COMP.	ICD9
Labbro	C00	escluso 9590-9989, 9050-9055, 9140	/3	140
Lingua	C01-02	escl. 9590-9989, 9050-9055, 9140	/3	141
Bocca	C03-06	escl. 9590-9989, 9050-9055, 9140	/3	143-145
Ghiandole salivari	C07-08	escl. 9590-9989, 9050-9055, 9140	/3	142
Orofaringe	C09-10	escl. 9590-9989, 9050-9055, 9140	/3	146
Rinofaringe	C11	escl. 9590-9989, 9050-9055, 9140	/3	147
Ipofaringe	C12-13	escl. 9590-9989, 9050-9055, 9140	/3	148
Faringe NAS	C14	escl. 9590-9989, 9050-9055, 9140	/3	149
Esofago	C15	escl. 9590-9989, 9050-9055, 9140	/3	150
Stomaco	C16	escl. 9590-9989, 9050-9055, 9140	/3	151
Intestino tenue	C17	escl. 9590-9989, 9050-9055, 9140	/3	152
Colon	C18	escl. 9590-9989, 9050-9055, 9140	/3	153
Retto	C19-21	escl. 9590-9989, 9050-9055, 9140	/3	154
Fegato	C22	escl. 9590-9989, 9050-9055, 9140	/3	155
Vie biliari	C23-24	escl. 9590-9989, 9050-9055, 9140	/3	156
Pancreas	C25	escl. 9590-9989, 9050-9055, 9140	/3	157
Cavità nasale	C30-31	escl. 9590-9989, 9050-9055, 9140	/3	160
Laringe	C32	escl. 9590-9989, 9050-9055, 9140	/3	161
Polmone	C33-34	escl. 9590-9989, 9050-9055, 9140	/3	162
Altri organi toracici	C37-38	escl. 9590-9989, 9050-9055, 9140	/3	163-164
Osso	C40-41	escl. 9590-9989, 9050-9055, 9140	/3	170
Pelle, melanomi	C44	8720-8790	/3	172
Pelle, non melanomi	C44	escl. 8720-8790, 9590-9989, 9050-9055, 9140	/3	173
Mesotelioma	tutte	9050-9055	/3	nd
Sarcoma di Kaposi	tutte	9140	/3	176
Tessuti molli	C47, 49	escl. 9590-9989, 9050-9055, 9140	/3	171
Mammella	C50	escl. 9590-9989, 9050-9055, 9140	/3	174-175
Utero, collo	C53	escl. 9590-9989, 9050-9055, 9140	/3	180
Utero, corpo	C54	escl. 9590-9989, 9050-9055, 9140	/3	182
Utero NAS	C55	escl. 9590-9989, 9050-9055, 9140	/3	179
Ovaio	C56	escl. 9590-9989, 9050-9055, 9140	/3	183.0
Altri genitali femminili	C51-52, 57, 58	escl. 9590-9989, 9050-9055, 9140	/3	181, 183.(2-9), 184
Pene	C60	escl. 9590-9989, 9050-9055, 9140	/3	187.(1-4)
Prostata	C61	escl. 9590-9989, 9050-9055, 9140	/3	185
Testicolo	C62	escl. 9590-9989, 9050-9055, 9140	/3	186
Altri genitali maschili	C63	escl. 9590-9989, 9050-9055, 9140	/3	187.(5-9)
Rene	C64	escl. 9590-9989, 9050-9055, 9140	/3	189.0
Vescica (maligno)	C67	escl. 9590-9989, 9050-9055, 9140	/3	188
Vescica (non maligno)	C67	escl. 9590-9989, 9050-9055, 9140	/1 /2	233.7, 236.7
Altre vie urinarie	C65-66, 68	escl. 9590-9989, 9050-9055, 9140	/3	189.(1-9)
Occhio	C69	escl. 9590-9989, 9050-9055, 9140	/3	190
Encefalo/SNC (maligno)	C70-72	escl. 9590-9989, 9050-9055, 9140	/3	191-192
Encefalo/SNC (non maligno)	C70-72	escl. 9590-9989, 9050-9055, 9140	/0 /1	225, 228.02, 237.(5-9)
Tiroide	C73	escl. 9590-9989, 9050-9055, 9140	/3	193
Altre ghiandole endocrine	C74-75	escl. 9590-9989, 9050-9055, 9140	/3	194
Linfoma di Hodgkin	tutte	9650-9667	/3	201
Linfoma non Hodgkin	tutte	9590-9596, 9670-9729, 9760-9764, 9940, 9948	/3	200, 202.(0-2, 4, 8, 9), 273.3
	escl. C42.0, 1, 4	9823, 9827		
Mieloma	tutte	9731-9734	/3	203, 238.6
Leucemia linfatica acuta	tutte	9826, 9835-9837	/3	204.0
	C42.0, 1, 4	9827		
Leucemia linfatica cronica	tutte	9831-9834	/3	204.1
	C42.0, 1, 4	9823		
Leucemia mieloide acuta	tutte	9840, 9861, 9866-9874, 9891-9920, 9931	/3	205.0, 206.0, 207.(0, 2)
Leucemia mieloide cronica	tutte	9863, 9875, 9876, 9945, 9946	/3	205.1
Altre MMPC e SMD	tutte	9950-9989	/3	207.1, 238.(4, 6, 7)
Leucemia NAS	tutte	9800-9820, 9860, 9930	/3	204-208 (escl. precedenti)
Miscellanea	tutte	9740-9758	/3	159.1, 202.(3, 5, 6)
	C42	escluso 9590-9989, 9050-9055, 9140		
Mal definite e metastasi	C26, C39, C48, C76-	escluso 9590-9989, 9050-9055, 9140	/3	158, 159.(0, 9), 165, 195-