

# MATERIALI E METODI

## Dati di incidenza, mortalità e popolazione

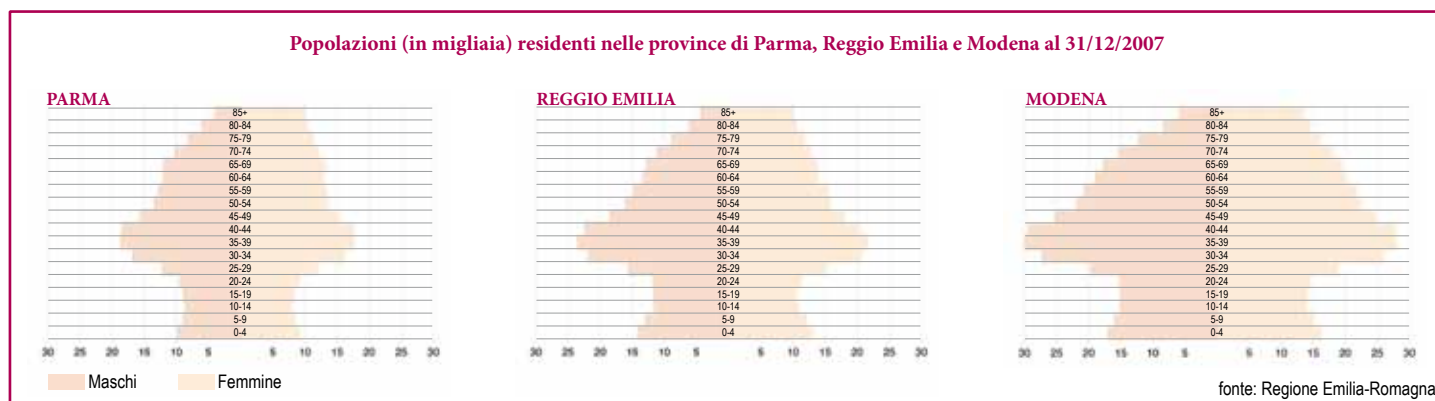
Nel presente lavoro sono considerati i casi incidenti (nuovi casi) di tumore maligno diagnosticati nel periodo 1996-2007 nelle popolazioni residenti nelle province emiliane di Parma, Reggio Emilia e Modena. Sono stati inclusi nell'incidenza anche i tumori della vescica in situ e a comportamento incerto, in base alle indicazioni dell'ENCR (*European Network for Cancer Registries*). Tali tumori sono stati, tuttavia, esclusi dalle analisi di sopravvivenza. Sono stati esclusi dal computo dell'incidenza i tumori non maligni del sistema nervoso centrale poiché le procedure di raccolta dati da parte dei registri per questo tipo di lesioni non sono ancora omogenee.

I casi incidenti di tumore sono selezionati in base alla Classificazione Internazionale per l'Oncologia, terza edizione (ICD-O-3). Topografia ed istologia dei tumori in analisi sono riportati

nell'allegato A.

Per l'analisi di mortalità sono considerati tutti i decessi per tumore che si sono verificati nelle popolazioni residenti nelle tre province negli anni 1996-2007. La mortalità è classificata in base alla Classificazione Internazionale delle Malattie, Traumatismi e Cause di morte ICD-IX revisione (ICD-9). L'allegato A presenta i raggruppamenti operati sui codici per l'identificazione dei decessi neoplastici.

La popolazione è rischio, utilizzata come denominatore nelle misure di occorrenza, è costituita dalla popolazione residente negli anni 1996-2007 nei comuni afferenti alle tre province (rilevata al 31/12 di ogni anno considerato). Di seguito sono riportate le piramidi d'età delle popolazioni nell'anno 2007 delle tre province.



## Tassi grezzi e tassi standardizzati

Il tasso grezzo di incidenza (e mortalità) è ottenuto come rapporto tra il numero di nuovi eventi che si verificano in un definito periodo di tempo e la popolazione a rischio di sperimentare l'evento durante tale periodo.

Indicando con  $O_j$  il numero di casi osservati, con  $j$  i gruppi di età (dove  $j = 1, \dots, 18$ ) e con  $n_j$  il numero di persone a rischio nel gruppo di età  $j$ -esimo, il tasso grezzo osservato nel gruppo di età  $j$ -esimo può essere scritto come  $T_j = O_j / n_j$ .

Il tasso grezzo pur indicativo dell'occorrenza del fenomeno, non permette, tuttavia, di confrontare due o più popolazioni in uno stesso arco temporale o una stessa popolazione a due o più tempi differenti. Questo perché il tasso grezzo dipende oltre che dall'intensità del fenomeno, anche dalla differente distribuzione per età delle popolazioni a confronto. Per effettuare correttamente i confronti è necessario ricorrere a una tecnica di aggiustamento chiamata standardizzazione diretta.

Il metodo della standardizzazione diretta consiste nell'applicare i tassi età-specifici dell'area o popolazione in studio ad una popolazione di riferimento assunta come standard ed esprime pertanto l'incidenza (o la mortalità) che sperimenterebbe la popolazione in studio se avesse l'incidenza (o mortalità) età-specifica della

popolazione standard.

Il tasso standardizzato con questo metodo in sostanza consiste in una media ponderata dei tassi specifici per età ottenuta usando come pesi la struttura per età della popolazione assunta come standard (Rotman 2002), dove i pesi corrispondono al numero di persone a rischio in ciascun gruppo di età della popolazione standard. Il tasso standardizzato con metodo diretto è espresso pertanto come:

$$T_{STD} = \frac{\sum_j w_j T_j}{\sum_j w_j}$$

dove  $T_j$  è il tasso nella classe d'età  $j$ -esima e  $w_j$  è il peso per la classe d'età  $j$ -esima che deriva dalla popolazione standard. L'errore standard del tasso standardizzato è (Cochran 1977):

$$ES(T_{STD}) = \frac{1}{\sum w_j} \sqrt{\sum_j \frac{w_j^2 T_j (I - T_j)}{n_j}}$$

L'intervallo di confidenza al 95% è quindi calcolato come:  $T_{STD} \pm 1.96 ES(T_{STD})$

Il tasso d'incidenza è espresso come numero di casi per 100000 residenti.

In questa pubblicazione è stata principalmente utilizzata come standard la popolazione europea (Waterhouse et al. 1976). Nelle tabelle 3 e 5 della scheda "Tutti i tumori" è proposta anche la standardizzazione, per le sedi principali in analisi, alla popolazione italiana al censimento 2001 (maschi e femmine *combined*) e alla popolazione mondiale (Segi 1960).

I tassi sono stati calcolati utilizzando il software Stata (StataCorp LP, College Station, TX, USA).

### Rischio cumulativo

Il rischio cumulativo rappresenta la probabilità per un individuo di sviluppare la malattia in studio (o di morire per la stessa) durante un certo intervallo temporale in assenza di altre cause di morte.

Nel presente rapporto sono stati calcolati i rischi cumulativi di ammalarsi o di morire per un certo tumore, in diversi periodi d'età della vita: 0-54, 0-64, 0-74 e 0-84 anni.

Il rischio cumulativo è calcolato nel seguente modo:

$$\text{Rischio cumulativo} = 1000 [1 - e^{-5(\sum T_j 10^3)}]$$

dove  $T_j$  rappresenta il tasso specifico per classe quinquennale d'età.

### Stima della variazione annua percentuale del tasso standardizzato (APC)

I trend temporali per l'incidenza e la mortalità sono analizzati confrontando i tassi annuali standardizzati per età alla popolazione europea. La stima della variazione annua percentuale (o *Annual Percentage Change - APC*) dei tassi è calcolata utilizzando il software statistico *Joinpoint* (versione 3.3.1) sviluppato dal *National Cancer Institute*. L'analisi *joinpoint* evidenzia un eventuale cambiamento in una serie storica di valori, sintetizzandoli in una spezzata ossia una serie di rette con pendenze diverse e collegate tra loro da punti detti nodi (*joinpoint*). In specifico, essa è calcolata mediante una regressione lineare in cui il logaritmo naturale del tasso rappresenta la variabile dipendente e l'anno di calendario è la variabile indipendente, cioè:  $y = mx + b$  dove,  $y = \ln(T_{STD})$  ed  $x$  rappresenta l'anno di calendario. L'APC è allora dato da:  $100(e^m - 1)$ .

Testare l'ipotesi che l'APC è uguale a zero è equivalente a testare l'ipotesi che il parametro di regressione dell'anno di calendario ( $m$ ) è uguale a zero. Sono stati considerati significativi i trend aventi associato un *p-value* minore al cut-off di  $\alpha=0.05$ . Il test statistico è effettuato mediante il metodo delle permutazioni Monte Carlo ( $n=4000$ ), (Kim et al. 2000).

### Stima del rischio relativo (RR) per comune per il totale dei tumori

Al fine della rappresentazione in mappe del rischio di tumore nei comuni afferenti alle tre province, è stato calcolato il rapporto standardizzato di incidenza SIR (*Standardized Incidence Ratio*). Esso rappresenta una stima del rischio relativo (RR) ed evidenzia quanto una popolazione (area) in esame sperimenta l'evento in studio (incidenza di neoplasia) in più (o in meno) rispetto ad una popolazione standard al netto dell'effetto dell'età. Il SIR è calcolato come rapporto tra il numero di casi osservati, in una popolazione in esame in un dato arco temporale, ed il numero di casi attesi nella popolazione se questa avesse lo stesso tasso d'incidenza età e sesso specifico di una popolazione standard di riferimento. La determinazione dei casi attesi è stata effettuata con

standardizzazione interna (Banerjee et al. 2004), usando come riferimento il tasso d'incidenza medio, età-specifico nel periodo 1996-2007 delle tre province.

La rappresentazione in mappe geografiche di indicatori di rischio ha l'obiettivo di facilitare la lettura dei dati, evidenziare la variabilità nella struttura spaziale di un fenomeno ed, eventualmente, mettere in evidenza aggregazioni di aree a maggior rischio. La semplice rappresentazione in mappa del RR stimato dal SIR presenta, tuttavia, alcuni svantaggi (Lawson et al. 2000): i) una potenziale instabilità della mappa che deriva dal fatto che la varianza del rischio è tanto più elevata in un'area quanto più è piccolo il numero di popolazione a rischio e, viceversa, tanto più piccola quanto la numerosità della popolazione è alta; ii) nessuna differenziazione tra le aree in assenza di casi; iii) nessun tentativo di cogliere la struttura spaziale del fenomeno. Per ovviare al problema è possibile fare ricorso a diverse soluzioni statistiche; tra queste, gli stimatori bayesiani gerarchici, i quali permettono di esprimere stime *smoothed* ("lisciate") del rischio. Attraverso questa metodologia è infatti possibile ridurre il "rumore" casuale e stabilizzare la stima locale attraverso una ponderazione del rischio relativo del singolo comune che sfrutta l'informazione derivante dai comuni adiacenti. È stato qui utilizzato un approccio completamente bayesiano, con ricorso a un modello condizionale autoregressivo (CAR). Nello specifico, il modello di riferimento è quello gerarchico di convoluzione gaussiana proposto da Besag et al. (1991). Il modello per la stima del RR di ciascuna area comunale è formulato come segue:

$$O_i \sim \text{Poisson}(E_i \lambda_i)$$

dove  $E_i$  sono i casi attesi nell'area  $i$ -esima e  $\lambda_i$  è il rischio relativo nell'area  $i$

$$\log(\lambda_i) = a + v_i + u_i$$

ossia a ciascun comune è assegnato una distribuzione a priori log-normale, in cui il parametro del rischio relativo è funzione lineare di un valore comune ( $a$ : intercetta), una componente eterogenea che non dipende dalla locazione geografica ( $v_i$ ) ed una componente autocorrelata che riflette la struttura spaziale dei comuni della provincia, incorporando l'influenza delle unità geografiche adiacenti ( $u_i$ ).

Le distribuzioni a posteriori sono ottenute per simulazione MCMC (metodo Monte Carlo per le catene di Markov) utilizzando il software Winbugs 1.4 (100000 iterazioni di cui 50000 *burn-in*).

### Sopravvivenza

L'analisi di sopravvivenza è stata sviluppata includendo i primi tumori maligni (cioè a comportamento /3 in base alla classificazione ICD-O-3) diagnosticati dal 1996 al 2007, ed escludendo i tumori della cute e quelli notificati in base all'esclusivo certificato di decesso (DCO). Il follow-up è stato chiuso alla data del 31/12/2008. La sopravvivenza (osservata e relativa), rispettivamente a uno, tre e cinque anni dalla diagnosi è stata stimata per le principali sedi tumorali.

Per tutte le analisi di sopravvivenza è stato utilizzato il software Stata (StataCorp LP, College Station, TX, USA).

La sopravvivenza osservata è una stima della probabilità di sopravvivenza per tutte le cause di morte in uno specifico intervallo temporale nei casi di tumore.

La sopravvivenza relativa (SR), invece, rappresenta una stima

della sopravvivenza per tumore in assenza di altre cause di morte.

La SR è, infatti, calcolata come rapporto fra la sopravvivenza osservata e la sopravvivenza attesa di un gruppo della popolazione generale con caratteristiche comparabili ai pazienti in studio. Nel presente rapporto per il calcolo della sopravvivenza attesa sono state utilizzate le tavole di sopravvivenza della Regione Emilia Romagna.

Sia la sopravvivenza osservata che quella relativa sono state stimate con il metodo di Hakulinen (Hakulinen 1982) e utilizzando il cosiddetto approccio completo.

Al fine di facilitare il confronto tra la SR stimata in diversi periodi di diagnosi (1996-99; 2000-03; 2004-07) e, quando possibile, per sottosede o stadio, la SR è stata standardizzata per età. La stima è stata effettuata applicando le classi di età ed i pesi proposti da Corazziari et al. (2004) e utilizzando il metodo suggerito da Brenner et al. (2004).

Al fine di mettere in luce differenze fra le SR fra periodi di diagnosi e sesso in termini di rischio relativo, è stata utilizzata la regressione di Poisson, come proposto da Dickman et al. (2004).

## BIBLIOGRAFIA

Banerjee, S., Carlin, B.P., Gelfand, A.E. (2004). *Hierarchical Modelling and Analysis for Spatial Data*. Chapman and Hall/CRC Press, Boca Raton, Florida.

Besag, J., York, J., Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics* 43: 1-59.

Brenner, H., Arndt, V., Gefeller, O., Hakulinen, T. (2004). An alternative approach to age adjustment of cancer survival rates. *European Journal of Cancer* 40(15): 2317-2322.

Cochran, W.G. (1977). *Sampling Techniques*. 3d ed. John Wiley & Sons, New York.

Corazziari, I., Quinn, M., Capocaccia, R. (2004). Standard cancer patient population for age standardising survival ratios. *European Journal of Cancer* 40(15): 2307-16.

Dickman, P.W., Sloggett, A., Hills, M., Hakulinen, T. (2004). Regression models for relative survival. *Statistics in Medicine* 23: 51-64.

Hakulinen, T. (1982). Cancer Survival Corrected for Heterogeneity in Patient Withdrawal. *Biometrics* 38: 933-942.

Kim, H.-J., Fay, M.P., Feuer, E.J., Midthune, D.N. (2000). Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine* 19: 335-351.

Lawson, A.B., Biggeri, A., Boehning, D., Lesaffre, E., Viel, J.F., Clark, A., Schlattmann, P., Divino, F. (2000). Disease mapping models: an empirical evaluation. *Statistics in Medicine* 19: 2217-2242.

Rothman, K.J. (2002). *Epidemiology: An Introduction*. Oxford University Press, New York.

Segi, M. (1960). *Cancer mortality for selected sites in 24 countries (1950-57)*. Department of Public Health, Tabokyo Univ. School of Medicine. Soudai, Japan.

Waterhouse, J., Muir, C., Correa, P., Powel, J. (eds) (1976). *Cancer Incidence in Five Continents, Vol. III*. IARC Scientific Publication No. 15. International Agency for Research on Cancer, World Health Organization, Lyon.

**Allegato A: Codifica dei tumori in base ai sistemi di classificazione ICD-O-3 e ICD-9**

Descrizione	CODICI INCIDENZA		CODICI MORTALITA'
	Topografia (ICD-O-3)	Istologia (ICD-O-3)*	ICD-9
Labbro	C00.0-C00.9	escluso 9590-9989, 9050-9055, 9140	140
Lingua	C01.9-C02.9	escluso 9590-9989, 9050-9055, 9140	141
Bocca	C03.0-C03.9, C04.0-C04.9, C05.0-C05.9, C06.0-C06.9	escluso 9590-9989, 9050-9055, 9140	143, 144, 145
Ghiandole salivari	C07.9-C08.9	escluso 9590-9989, 9050-9055, 9140	142
Orofaringe	C09.0-C09.9, C10.0-C10.9	escluso 9590-9989, 9050-9055, 9140	146
Rinofaringe	C11.0-C11.9	escluso 9590-9989, 9050-9055, 9140	147
Ipopfaringe	C12.9, C13.0-C13.9	escluso 9590-9989, 9050-9055, 9140	148
Faringe NAS	C14.0, C14.2-C14.8	escluso 9590-9989, 9050-9055, 9140	149
Esofago	C15.0-C15.9	escluso 9590-9989, 9050-9055, 9140	150
Stomaco	C16.0-C16.9	escluso 9590-9989, 9050-9055, 9140	151
Intestino tenue	C17.0-C17.9	escluso 9590-9989, 9050-9055, 9140	152
Colon	C18.0-C18.9	escluso 9590-9989, 9050-9055, 9140	153
Retto	C19.9, C20.9, C21.0-C21.2, C21.8	escluso 9590-9989, 9050-9055, 9140	154
Fegato e dotti biliari intraepatici	C22.0, C22.1	escluso 9590-9989, 9050-9055, 9140	155
Colecisti e vie biliari	C23.9, C24.0-C24.9	escluso 9590-9989, 9050-9055, 9140	156
Pancreas	C25.0-C25.9	escluso 9590-9989, 9050-9055, 9140	157
Cavità nasale	C30.0-C30.1, C31.0-C31.9	escluso 9590-9989, 9050-9055, 9140	160
Laringe	C32.0-C32.9	escluso 9590-9989, 9050-9055, 9140	161
Polmone, bronchi e trachea	C34.0-C34.9, C33.9	escluso 9590-9989, 9050-9055, 9140	162
Altri organi toracici	C37.9, C38.0-C38.8	escluso 9590-9989, 9050-9055, 9140	163, 164
Osso e cartilagine articolare	C40.0-C41.9	escluso 9590-9989, 9050-9055, 9140	170
Melanoma cutaneo	C44.0-C44.9	8720-8790	172
Cute (escluso melanoma)	C44.0-C44.9	escluso 8720-8790, 9590-9989, 9050-9055, 9140	173
Mesotelioma	Tutte le sedi	9050-9055	-
Sarcoma di Kaposi	Tutte le sedi	9140	-
Tessuti molli	C47.0-C47.9, C49.0-C49.9	escluso 9590-9989, 9050-9055, 9140	171
Mammella	C50.0-C50.9	escluso 9590-9989, 9050-9055, 9140	174-175
Cervice dell'utero	C53.0-C53.9	escluso 9590-9989, 9050-9055, 9140	180
Corpo dell'utero	C54.0-C54.9	escluso 9590-9989, 9050-9055, 9140	182
Utero NAS	C55.9	escluso 9590-9989, 9050-9055, 9140	179
Ovaio	C56.9	escluso 9590-9989, 9050-9055, 9140	183
Altri genitali femminili	C51.0-C51.9, C52.9, C57.0-C58.9	escluso 9590-9989, 9050-9055, 9140	181, 184
Pene	C60.0-C60.9	escluso 9590-9989, 9050-9055, 9140	187.1-187.4
Prostata	C61.9	escluso 9590-9989, 9050-9055, 9140	185
Testicolo	C62.0-C62.9	escluso 9590-9989, 9050-9055, 9140	186
Altri genitali maschili	C63.0-C63.9	escluso 9590-9989, 9050-9055, 9140	187.5-187.9 **
Rene, vie urinarie	C64.9, C65.9, C66.9, C68.0-C68.9	escluso 9590-9989, 9050-9055, 9140	189
Vescica	C67.0-C67.9	escluso 9590-9989, 9050-9055, 9140	188
Occhio	C69.0-C69.9	escluso 9590-9989, 9050-9055, 9140	190
Encefalo e altro SNC	C70.0-C70.9, C71.0-C71.9, C72.0-C72.9	escluso 9590-9989, 9050-9055, 9140	191-192
Tiroide	C73.9	escluso 9590-9989, 9050-9055, 9140	193
Altre ghiandole endocrine	C74.0-C74.9, C75.0-C75.9	escluso 9590-9989, 9050-9055, 9140	194
Linfoma di Hodgkin	Tutte le sedi	9650-9667	201
Linfoma non Hodgkin	C02.4, C09.8-C09.9, C11.1, C14.2, C37.9, C42.2, C77.0-C77.9	9590-9596, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687, 9689-9691, 9695, 9698-9702, 9705, 9708-9709, 9714-9719, 9727-9729, 9823, 9827	200, 202.0-202.2, 202.8-202.9
	Tutte le sedi eccetto: C02.4, C09.8-C09.9, C11.1, C14.2, C37.9, C42.2, C77.0-C77.9	9590-9596, 9670-9671, 9673, 9675, 9678-9680, 9684, 9687, 9689-9691, 9695, 9698-9702, 9705, 9708-9709, 9714-9719, 9727-9729	
	Tutte le sedi eccetto: C02.4, C09.8-C09.9, C11.1, C14.2, C37.9, C42.0-C42.2, C42.4, C77.0-C77.9	9823, 9827	
Mieloma	Tutte le sedi	9731-9732, 9734	203.0, 238.6 **
Leucemia linfatica acuta	Tutte le sedi	9826, 9835-9837	204.00.00
Leucemia linfatica cronica	C42.0, C42.1, C42.4	9823	204.1
Altre leucemie linfatiche	Tutte le sedi	9820, 9832-9834, 9940	202.4, 204.2, 204.8-204.9 **
Leucemia mieloide acuta	Tutte le sedi	9840, 9861, 9866, 9867, 9871-9874, 9895-9897, 9910, 9920	205.0, 207.0, 207.2
Leucemia mieloide cronica	Tutte le sedi	9863, 9875, 9876, 9945, 9946	205.01.00
Leucemia monocitica acuta	Tutte le sedi	9891	206.00.00
Altre leucemie mieloidi/monocitiche	Tutte le sedi	9860, 9930	205.2-205.3, 205.8-205.9, 206.1-206.2, 206.8-206.9 **
Altre leucemie e leucemie NAS	Tutte le sedi	9733, 9742, 9800, 9801, 9805, 9831, 9870, 9931, 9948	203.1, 207.8, 208.0-208.2, 208.8-208.9
	C42.0, C42.1, C42.4	9827	
MMPC e SMD	Tutte le sedi	9950-9989	207.1, 238.4, 238.7
Miscellanea	Tutte le sedi	9740-9741, 9750-9758, 9760-9769	159.1, 202.3, 202.5, 202.6, 203.8
	C42.0-C42.4	escluso 9590-9989, 9050-9055, 9140	
Mal definite e metastasi	C26.0-C26.9, C39.0-C39.9, C48.0-C48.8, C76.0-C76.8, C77.0-C77.9, C80.9	escluso 9590-9989, 9050-9055, 9140	158, 159.0, 159.8-159.9, 165, 195-199

\* Sono considerati incidenti i tumori con codice di comportamento /3, con l'eccezione dei tumori della vescica. Per questa sede anche i tumori in situ (/2) e a comportamento incerto (/1) sono inclusi nell'incidenza, anche se esclusi dall'analisi di sopravvivenza. \*\* Sono inclusi, per gli anni precedenti al 2007, anche i codici senza specifica della sottosede, rispettivamente: 187 in "Altri genitali maschili"; 203 in "Mieloma"; 204 in "Altre leucemie linfatiche"; 205 in "Altre leucemie mieloidi/monocitiche". - Non disponibile in base alla classificazione ICD-9.