

La banca dati dei registri tumori italiani: la stima delle popolazioni osservate

The database of the italian cancer registries: estimates of the observed populations

Stefano Guzzinati,¹ Alessandra Spitale,² Guido Miccinesi,³ Paola Zambon,⁴ Stefano Rosso²

¹Registro tumori del Veneto, Azienda ospedaliera di Padova

²Registro tumori Piemonte, Centro per l'epidemiologia e la prevenzione oncologica (CPO)

³Registro tumori toscano, Centro per lo studio e la prevenzione oncologica (CSPO)

⁴Registro tumori del Veneto, Università degli studi di Padova

Corrispondenza: Stefano Guzzinati, Registro tumori del Veneto, via Gattamelata 64, 35128 Padova; tel. 049-8076412; fax 049-8076789; e-mail: stefano.guzzinati@unipd.it

Cosa si sapeva già

■ Le popolazioni osservate dai registri tumori (RT) italiani erano disponibili a un livello di dettaglio non sufficiente per l'analisi dei *trend* di incidenza e mortalità; per esempio di alcuni RT si conosceva solo la distribuzione per classi quinquennali d'età e non per classi annuali.

Cosa si aggiunge di nuovo

■ È stato creato e reso disponibile un archivio con le stime delle popolazioni osservate dagli RT italiani con il dettaglio massimo ottenibile: suddivisione per sesso, classe annuale d'età e anno di calendario.
■ Utilizzando diversi metodi di interpolazione si rende disponibile un foglio Excel che permette di inserire una popolazione per classi quinquennali e ottenere una stima delle classi annuali.

Riassunto

Obiettivi: costruire un archivio con la stima delle popolazioni osservate dai registri tumori (RT) italiani, suddivise per sesso, classe annuale d'età e anno di calendario; rendere disponibile tale archivio nel sito dell'AIRT; rendere disponibile un foglio Excel che permette di inserire una popolazione per classi quinquennali e ottenere una stima delle classi annuali.

Disegno: utilizzazione di diversi metodi di interpolazione, quali stima della popolazione per anni intercensuari, stima della popolazione per classi annuali d'età partendo da classi quinquennali, stima della classe finale aperta di 85 anni e più.

Setting: AIRT.

Partecipanti: 17 RT partecipanti allo studio dei *trend* di incidenza e mortalità promosso dall'AIRT nel 2003.

Outcome principali: archivio delle stime delle popolazioni osservate dagli RT per sesso e classi annuali d'età al 31 dicembre di ogni anno di incidenza.

Risultati: soltanto sei su 17 RT che partecipano allo studio AIRT dei *trend* disponevano dei dati richiesti; per gli altri RT sono stati utilizzati diversi metodi di interpolazione, la cui *performance*, saggiata con una misura di errore empirica detta WMISE, ha portato a buoni risultati, anche se nelle classi anziane l'adattamento è minore e in alcuni casi insufficiente.

Conclusioni: sono stati utilizzati diversi metodi di interpolazione per costituire e rendere disponibile un archivio contenente stime affidabili delle popolazioni osservate, da utilizzare come denominatori per l'analisi dei *trend* di incidenza e mortalità.

Si rende altresì disponibile un foglio Excel che permette di inserire una popolazione per classi quinquennali d'età e ottenere la stima per classi annuali.

(*Epidemiol Prev* 2004; 28 (4-5): 231-38)

Parole chiave: archivio, interpolazione, popolazione

Abstract

Objectives: building a database with the estimates of the observed populations of the Italian Cancer Registries (CR), subdivided into sex, annual age classes and calendar year; making this population database available in the AIRT website; making available an Excel file that allows the input of a population by 5-year age classes in order to obtain an estimate of the annual classes.

Design: employment of different interpolation methods, namely a population estimate of intercensal years, a population estimate of

annual age classes starting from 5-year classes, an estimate of the final open class of 85 years and over.

Participants: 17 CRs that take part to the incidence and mortality trend study promoted by AIRT in 2003.

Setting: AIRT.

Main Outcome: a database of the estimated populations observed by the CRs by sex and annual age classes on 31 December of each incidence year.

Results: only for 6 on 17 CRs taking part in the AIRT trend study we have the requested data; for the other CRs different in-

terpolation methods have been used. The performance of these methods, tested with an empirical measure of error called WMI-SE, have given good results, even if in the elderly classes the fit is poor and sometimes not sufficient.

Conclusions: different interpolation methods have been used to build and make available a database containing reliable data of population, in order to use them as denominators for the

incidence and mortality trend analysis. It is also available an Excel file that allows the input of a population by 5-year age classes in order to obtain an esteem of the annual classes. (*Epidemiol Prev* 2004; 28 (4-5): 231-38)

Key words: database, interpolation, population

Introduzione

Obiettivo del presente lavoro è la realizzazione di un archivio informatizzato contenente la stima delle popolazioni osservate dai registri tumori, attivi nel nostro paese e accreditati dall'Associazione italiana registri tumori (AIRT), secondo il massimo dettaglio possibile di informazione.

Il calcolo di indicatori di occorrenza delle patologie richiede di poter disporre non solo di numeratori completi e adeguati, ma anche di corrispondenti denominatori, ovvero di una

stima corretta della popolazione a rischio. Fra le numerose istituzioni che, a vario titolo e con diversi obiettivi, fanno uso di dati relativi alla numerosità delle popolazioni, i registri tumori (RT) dispongono dei dati relativi alle popolazioni osservate secondo le rilevazioni censuarie e annuali ISTAT o direttamente dalle anagrafi comunali.

La disponibilità dei dati relativi alla popolazione residente coperta dagli RT ha rappresentato un problema fino al 1990; la rilevazione annuale ISTAT sulla popolazione residente nei comuni per sesso, anno di nascita e stato civile al 31 dicembre (denominata «POSAS») è stata istituita soltanto dal 1992. La copertura della rilevazione è andata sensibilmente crescendo, passando dal 77% alla data del 31.12.1992 al 98% al 31.12.1997.¹

Per gli anni precedenti e per le zone del territorio nazionale non coperte dalla rilevazione dell'ISTAT si dispone, solitamente, di dati aggregati per classi d'età e periodi. Tali aggregazioni, tuttavia, spesso non soddisfano l'esigenza del ricercatore, che deve calcolare indicatori statistici su classi di età e periodi non previsti e quindi non disponibili direttamente. Un esempio di tale evenienza è offerto dallo studio sugli andamenti temporali di incidenza e mortalità per tumori in Italia, promosso dall'AIRT, per cui dovevano essere calcolati i tassi specifici per sesso, età e anno di calendario.² Data la non omogeneità dell'informazione disponibile è stato quindi necessario procedere alla stima di alcuni disaggregati di popolazione, fornendo per ogni registro la distribuzione della popolazione residente per anno di calendario, sesso e classe annuale d'età. Le tecniche di stima a disposizione sono numerose, come pure le peculiarità delle distribuzioni di popolazioni, attraversate da differenti eventi demografici che ne hanno inciso profili disomogenei. Tali tecniche sono state sviluppate prevalentemente in ambito demografico e ampiamente discusse in quella letteratura scientifica.³

In questo lavoro vengono utilizzate diverse tecniche di stima delle popolazioni per affrontare e risolvere le difficoltà che sorgono per la definizione della banca dati delle popolazioni osservate dai registri tumori italiani a partire da informazioni disomogenee per i vari registri. Ciò non esaurisce naturalmente le diverse scelte metodologiche di stima, ma considera solo quelle ritenute più idonee a partire dai dati a disposizione. Inoltre si intende condividere con altri possibili utilizzatori la nostra esperienza per estendere le stime anche ad altre popolazioni.

Sottogruppo interpolato	G1	G2	G3	G4	G5
1° pannello					
1° quinto di G1	0,3333	-0,1636	-0,0210	0,0796	-0,0283
2° quinto di G1	0,2595	-0,0780	0,0130	0,0100	-0,0045
3° quinto di G1	0,1924	0,0064	0,0184	-0,0256	0,0084
4° quinto di G1	0,1329	0,0844	0,0054	-0,0356	0,0129
5° quinto di G1	0,0819	0,1508	-0,0158	-0,0284	0,0115
2° pannello					
1° quinto di G2	0,0404	0,2000	-0,0344	-0,0128	0,0068
2° quinto di G2	0,0093	0,2268	-0,0402	0,0028	0,0013
3° quinto di G2	-0,0108	0,2272	-0,0248	0,0112	-0,0028
4° quinto di G2	-0,0198	0,1992	0,0172	0,0072	-0,0038
5° quinto di G2	-0,0191	0,1468	0,0822	-0,0084	-0,0015
pannello centrale					
1° quinto di G3	-0,0117	0,0804	0,1570	-0,0284	0,0027
2° quinto di G3	-0,0020	0,0160	0,2200	-0,0400	0,0060
3° quinto di G3	0,0050	-0,0280	0,2460	-0,0280	0,0050
4° quinto di G3	0,0060	-0,0400	0,2200	0,0160	-0,0020
5° quinto di G3	0,0027	-0,0284	0,1570	0,0804	-0,0117
penultimo pannello					
1° quinto di G4	-0,0015	-0,0084	0,0822	0,1468	-0,0191
2° quinto di G4	-0,0038	0,0072	0,0172	0,1992	-0,0198
3° quinto di G4	-0,0028	0,0112	-0,0248	0,2272	-0,0108
4° quinto di G4	0,0013	0,0028	-0,0402	0,2268	0,0093
5° quinto di G4	0,0068	-0,0128	-0,0344	0,2000	0,0404
ultimo pannello					
1° quinto di G5	0,0115	-0,0284	-0,0158	0,1508	0,0819
2° quinto di G5	0,0129	-0,0356	0,0054	0,0844	0,1329
3° quinto di G5	0,0084	-0,0256	0,0184	0,0064	0,1924
4° quinto di G5	-0,0045	0,0100	0,0130	-0,0780	0,2595
5° quinto di G5	-0,0283	0,0796	-0,0210	-0,1636	0,3333

Tabella 1. Coefficienti di interpolazione basati sulla formula «ordinaria» di Beers.

Table 1. Interpolation coefficients based on the Beers «ordinary» formula.

Registro tumori	Periodo
Genova	1993-1996
Latina	1988-1991
Macerata	1991-1997
Napoli	1996-1998
Ragusa	1994-1997
Romagna	1986-1992
Sassari	1992-1997
Varese	1976-1997

Tabella 2. Registri tumori per cui è stato applicato il metodo di Beers per la stima della popolazione nelle classi annuali.

Table 2. Cancer Registries in which the Beers method was applied for the population evaluation of 1-year age classes.

Materiali e metodi

I metodi di stima si avvalgono di tecniche d'interpolazione utilizzate anche in ambito non strettamente demografico. In generale, quando conosciamo una serie finita di misure sperimentali o osservazioni, la stima di un valore ignoto può essere calcolata utilizzando i dati disponibili attraverso un'operazione detta d'interpolazione.

L'interpolazione può essere utilizzata per esempio per:

- ottenere una stima della popolazione negli anni intercensuali;
- suddividere dati raggruppati in singole componenti, come stimare classi annuali d'età partendo dalle classi quinquennali;
- suddividere una classe di età aperta (per esempio 75 e più anni) in più classi utili all'analisi (per esempio 75-79, 80-84, 85 e più anni).

Nel caso di fluttuazioni nella crescita o diminuzione di una popolazione, o nella distribuzione nelle classi d'età, dovute a eventi quali guerre, migrazioni, epidemie, eccetera, si impiegano tecniche di «lisciamiento» dei dati (*smoothing*) che, utilizzando delle interpolazioni, servono per eliminare valori estremi, in modo da identificare meglio *trend* e *pattern* reali.⁴

Nella fase di costruzione dell'archivio contenente le popolazioni osservate dagli RT suddivise per sesso, classe annuale d'età, anno di calendario, soltanto sei (Torino, Biella, Nord-Est, Firenze-Prato, Umbria) sui 17 RT partecipanti allo studio dei *trend* disponevano dei dati secondo i dettagli richiesti. Per i restanti RT sono stati utilizzati tre diversi metodi di stima.

Stima della popolazione per gli anni intercensuali

Partendo dalla distribuzione di una popolazione rilevata in due censimenti (a,b) si stima la popolazione (P) per l'anno intercensuale x, con la seguente formula di Waring:³

$$P(x) = P(a) \cdot (b-x)/(b-a) + P(b) \cdot (x-a)/(b-a)$$

Per esempio per i censimenti del 1981 e 1991, volendo stimare la popolazione dell'anno 1987, si procede come di seguito:

$$P(1987) = P('81) \cdot ('91-'87) / ('91-'81) + P('91) \cdot ('87-'81) / ('91-'81) = 0,4 P('81) + 0,6 P('91)$$

Questo metodo è stato utilizzato per stimare la popolazione dell'RT di Parma per gli anni 1982-86 e per l'RT del Veneto per il 1987-1990 partendo dai rispettivi dati dei censimenti del 1981 e del 1991.

Stima della popolazione nelle classi annuali d'età

Per stimare la popolazione nelle classi annuali d'età a partire dalle classi quinquennali sono disponibili una varietà di equazioni interpolanti,³ tra cui le più usate sono le formule di Sprague, Karup-King, Grabill, Beers (ordinaria e modificata).⁵ Tali formule possono essere espresse in forma di componenti lineari ossia di coefficienti applicati ai dati originari: un valore interpolato può essere facilmente calcolato moltiplicando i dati per i corrispondenti coefficienti e sommando i prodotti così ottenuti. In questo modo, come applicato anche in altri studi,⁶⁻⁷ si ha il vantaggio di «lisciare» i dati, smussando i valori estremi.

Nella tabella 1 sono presentati i coefficienti di interpolazione basati sulla formula ordinaria di Beers, usati per le stime delle classi annuali, essendo nota la distribuzione della popolazione per classi quinquennali.

Vengono utilizzati i multipli del pannello centrale per tutte le classi d'età, eccetto per la prima classe d'età 0-4 anni (1° pannello), per la classe 5-9 anni (2° pannello), per l'ultima classe d'età non aperta (ultimo pannello) e per la penultima classe (penultimo pannello).

Per esempio per stimare la popolazione nella classe d'età 34 anni (p_{34}) partendo da classi quinquennali, applichiamo i coefficienti relativi al 5° quinto del pannello centrale, essendo l'età di 34 anni il 5° quinto della classe 30-34:

$$p_{34} = (0,0027) p_{20-24} + (-0,0284) p_{25-29} + (0,1570) p_{30-34} + (0,0804) p_{35-39} + (-0,0117) p_{40-44}$$

dove p_{20-24} , p_{25-29} , p_{30-34} , p_{35-39} , p_{40-44} sono rispettivamente le popolazioni nelle classi d'età 20-24, 25-29, 30-34, 35-39 e 40-44 anni.

Il metodo di Beers è stato utilizzato per stimare le classi annuali partendo da quelle quinquennali per 8 RT (tabella 2).

Stima della classe d'età finale

Per la stima della classe d'età finale, che risultava molto disomogenea tra gli RT, andando dai 75 anni e più ai 110 e più, si è utilizzata un'interpolazione «pro-rata»:³ la classe finale è stata riportata a 85 e più anni per tutti i registri applicando la distribuzione percentuale della popolazione osservata nello stesso registro nell'anno più prossimo in cui era noto il dettaglio. Questa tecnica è stata impiegata per tre RT: Parma per il periodo 1982-1990, Modena per il 1988-1990 e Ragusa per il 1981-1997.

La *performance* dei metodi di stima e la bontà di adattamento della curva stimata rispetto alla osservata sono state studiate confrontando le stime con i dati di alcune popolazioni per cui era più facilmente reperibile il dettaglio per classi di età annuali. Si è dunque scelto Torino, Genova, Veneto, Parma e Firenze, Napoli e Ragusa, i cui territori sono coperti dal-

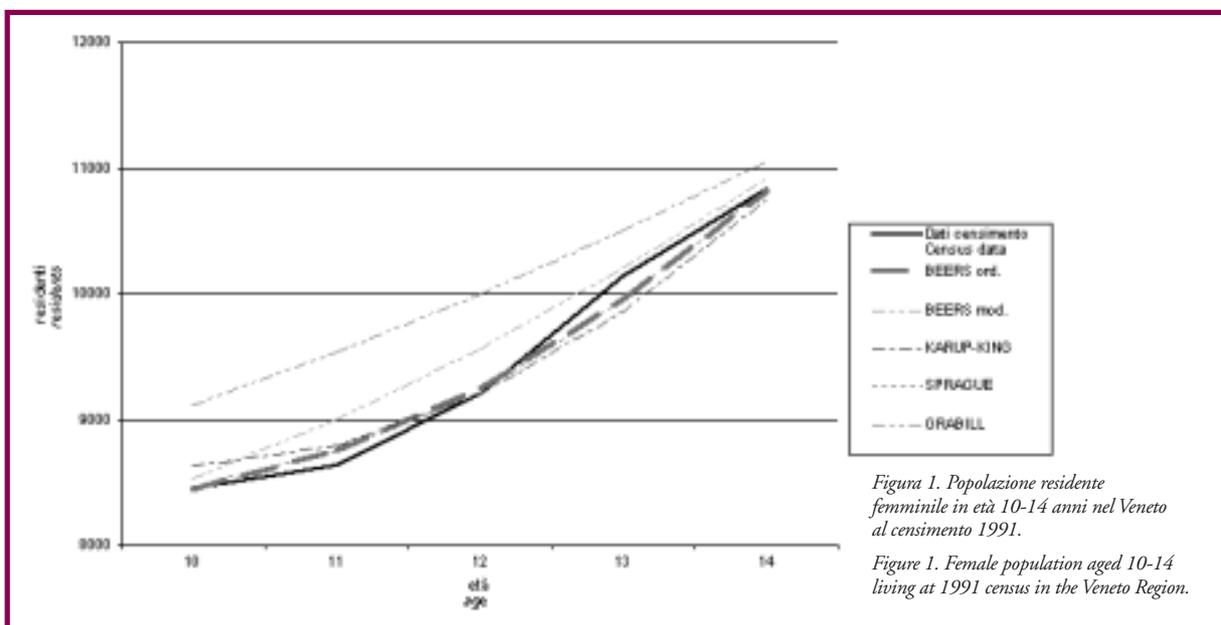


Figura 1. Popolazione residente femminile in età 10-14 anni nel Veneto al censimento 1991.

Figure 1. Female population aged 10-14 living at 1991 census in the Veneto Region.

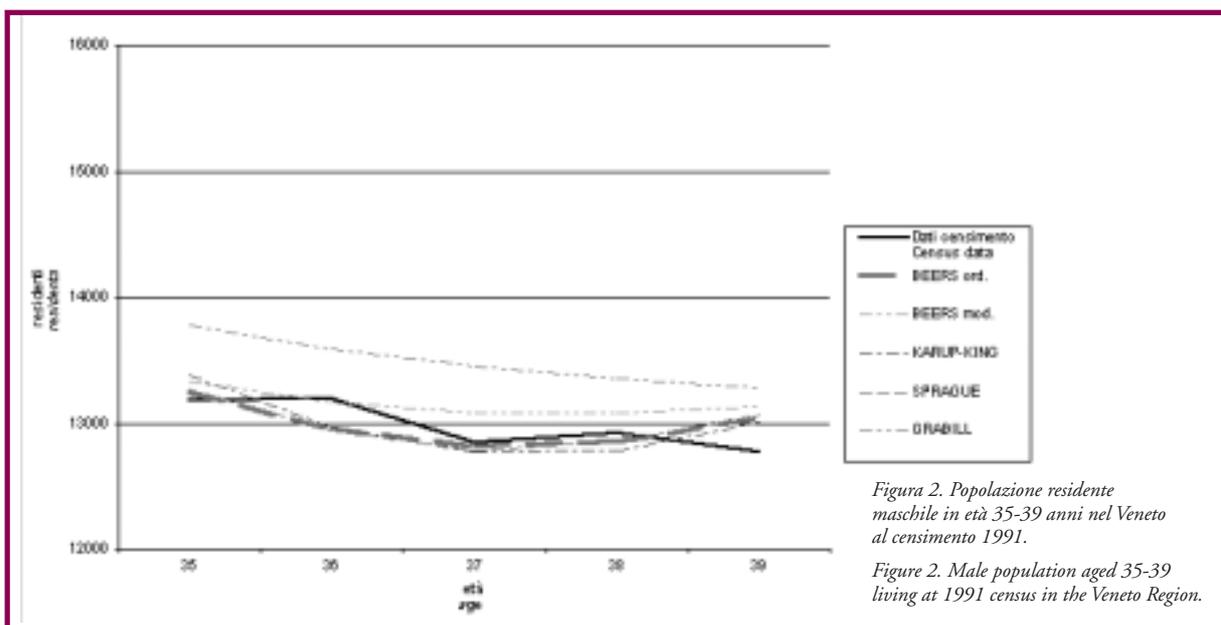


Figura 2. Popolazione residente maschile in età 35-39 anni nel Veneto al censimento 1991.

Figure 2. Male population aged 35-39 living at 1991 census in the Veneto Region.

l'attività dei registri tumori, in modo da rappresentare diverse realtà del Nord, Centro e Sud Italia. Per saggiare statisticamente la bontà d'adattamento delle interpolazioni ai dati non sono utilizzabili le tradizionali statistiche χ^2 per le frequenze, poiché la scala delle differenze (migliaia di osservazioni) rappresenta una dimensione in cui tali statistiche diventano troppo sensibili all'errore di primo tipo, generando risultati significativi in presenza di differenze irrilevanti (bassa specificità). Analogamente, trasformazioni di scala che consentono di accedere a test parametrici si rivelano, all'opposto, poco sensibili e potenti. Fortunatamente viene in soc-

corso una classe di statistiche utilizzate per le tecniche di *smoothing*; in particolare per le tecniche che utilizzano metodi non parametrici quali i *Kernel*.⁸ Inoltre, anche i metodi parametrici d'interpolazione lineare si possono riscrivere come casi particolari di una più generale funzione *Kernel*. Se infatti definiamo tale funzione come:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

UOMINI					
registro	età				totale
	0-14	15-44	45-64	65-84	
Torino	2,142	12,527	12,116	23,413	50,198
(p-value)	0,99995	0,99790	0,91203	0,32237	0,99928
Genova	3,291	12,320	9,430	16,222	41,264
(p-value)	0,99929	0,99820	0,97737	0,75704	0,99999
Veneto	1,373	4,930	9,390	28,529	44,222
(p-value)	1,00000	1,00000	0,97793	0,12579	0,99995
Parma	1,219	8,660	5,521	20,476	35,877
(p-value)	1,00000	0,99995	0,99941	0,49129	1,00000
Firenze - Prato	2,830	7,783	10,699	25,295	46,608
(p-value)	0,99972	0,99999	0,95366	0,23460	0,99984
Napoli	5,059	12,136	12,628	7,523	37,347
(p-value)	0,99161	0,99844	0,89276	0,99686	1,00000
Ragusa	1,380	13,080	7,456	23,327	45,243
(p-value)	1,00000	0,99687	0,99490	0,32679	0,99991
Gradi di libertà	15	30	20	21	86

DONNE					
registro	età				totale
	0-14	15-44	45-64	65-84	
Torino	1,155	5,505	8,366	25,673	40,699
(p-value)	1,00000	1,00000	0,98916	0,21919	0,99999
Genova	1,331	5,066	7,670	31,426	45,491
(p-value)	1,00000	1,00000	0,99384	0,06685	0,99990
Veneto	0,968	2,680	5,578	26,546	35,771
(p-value)	1,00000	1,00000	0,99936	0,18640	1,00000
Parma	1,546	1,957	3,094	46,779	53,376
(p-value)	0,99999	1,00000	0,99999	0,00101	0,99778
Firenze - Prato	1,560	5,291	7,422	38,231	52,503
(p-value)	0,99999	1,00000	0,99505	0,01211	0,99835
Napoli	2,553	7,988	13,550	17,252	41,343
(p-value)	0,99985	0,99998	0,85257	0,69575	0,99999
Ragusa	1,421	8,351	11,246	22,864	43,882
(p-value)	1,00000	0,99997	0,93958	0,35130	0,99995
Gradi di Libertà	15	30	20	21	86

Tabella 3. Statistiche di adattamento WMISE della curva stimata ai dati reali per classi d'età e sesso nelle popolazioni osservate in alcuni registri tumori (valore di p per una distribuzione χ^2 con gradi di libertà nell'ultima riga).

Table 3. WMISE statistics of the estimated curve to the real data by age classes and sex in the populations observed in some Tumour Registries (p-value for χ^2 distribution with degree of freedom in the last line)

dove K è una funzione che definisce dei coefficienti a seconda del valore che assume $1/nh$, con h la finestra su cui agisce l'operatore K , possiamo definire tale operatore in modo che identifichi i coefficienti indicati dai vari metodi d'interpolazione precedentemente nominati (Beers, Sprague, Karup-King e Grabill). Le finestre su cui agisce l'operatore saranno poi le classi di età quinquennali di partenza, mentre la sommatoria opererà per ognuna delle classi di età annuali da definire.

Le misure d'errore per questa classe di metodi di *smoothing* non parametrici si basano sull'errore medio quadrato inte-

grato, o MISE (*mean integrated squared error*), dove l'integrazione è definita nello spazio campionario in base a una distribuzione conosciuta. Nel nostro caso, dalla formulazione generale del MISE:

$$MISE(h) = \int_{\mathcal{X}} \{E[\hat{f}_h(x)] - f(x)\}^2 dx + \int_{\mathcal{X}} \text{var}[\hat{f}_h(x)] dx$$

dove $\hat{f}_h(x)$ è lo stimatore di densità con finestra h , e $f(x)$ è la densità vera, si utilizza una formula approssimativa in cui le differenze fra stimato e osservato al quadrato sono moltiplicate per un peso proporzionale all'integrale di una distribuzione chi-quadro calcolata per il valore osservato: tale formulazione empirica prende il nome di *weighted mean integrated squared error* o WMISE.⁸

Per valutare poi quanto l'utilizzo di una popolazione «non esatta» incida nel calcolo dell'incidenza, sono stati infine confrontati i tassi di incidenza nel Veneto per il biennio 1995-96 calcolati sui dati del censimento 1991 con quelli calcolati utilizzando le stime della popolazione media 1995-96.⁹

Risultati

Soltanto sei sui 17 RT partecipanti allo studio dei *trend*² disponevano dei dati di popolazione secondo i dettagli richiesti, per gli altri RT sono state utilizzate varie tecniche di interpolazione.

In figura 1 e 2, a titolo esemplificativo, sono confrontate con i dati effettivi le stime ottenute con differenti metodi di interpolazione in un certo intervallo di età (10-14 anni e 35-39 anni). I risultati sono simili se si utilizzano i coefficienti di Beers (formula ordinaria), Sprague o Karup-King. Questi tre metodi forzano la somma delle popolazioni nelle cinque classi annuali al valore della classe quinquennale di partenza. Gli altri due metodi (Beers modificato e Grabill), che invece non hanno tale vincolo, si discostano maggiormente dai valori effettivi, pertanto si è scelto di utilizzare come metodo la formula ordinaria di Beers, utilizzata anche dal Center for Disease Control and Prevention per la costruzione di tavole di sopravvivenza.⁷

In figura 3 e 4 si mostrano i risultati relativi all'RT del Veneto, del tutto sovrapponibili a quanto osservato per gli altri RT saggiati (Torino e Genova). Il confronto mostra una buona corrispondenza tra valori stimati e osservati, con l'eccezione, per entrambi i sessi, delle età 45-46 anni e 72-75 anni corrispondenti a coorti di nati rispettivamente nel 1945-46 e nel 1916-19, periodo relativo alle due guerre mondiali, dove si osserva una certa irregolarità nei dati effettivi. Inoltre, si può notare una differenza sistematica per la popolazione entro il primo anno di vita. Questo è un problema generale dovuto da un lato alla più generale difficoltà dei metodi d'interpolazione nel lavorare efficientemente ai confini della distribu-

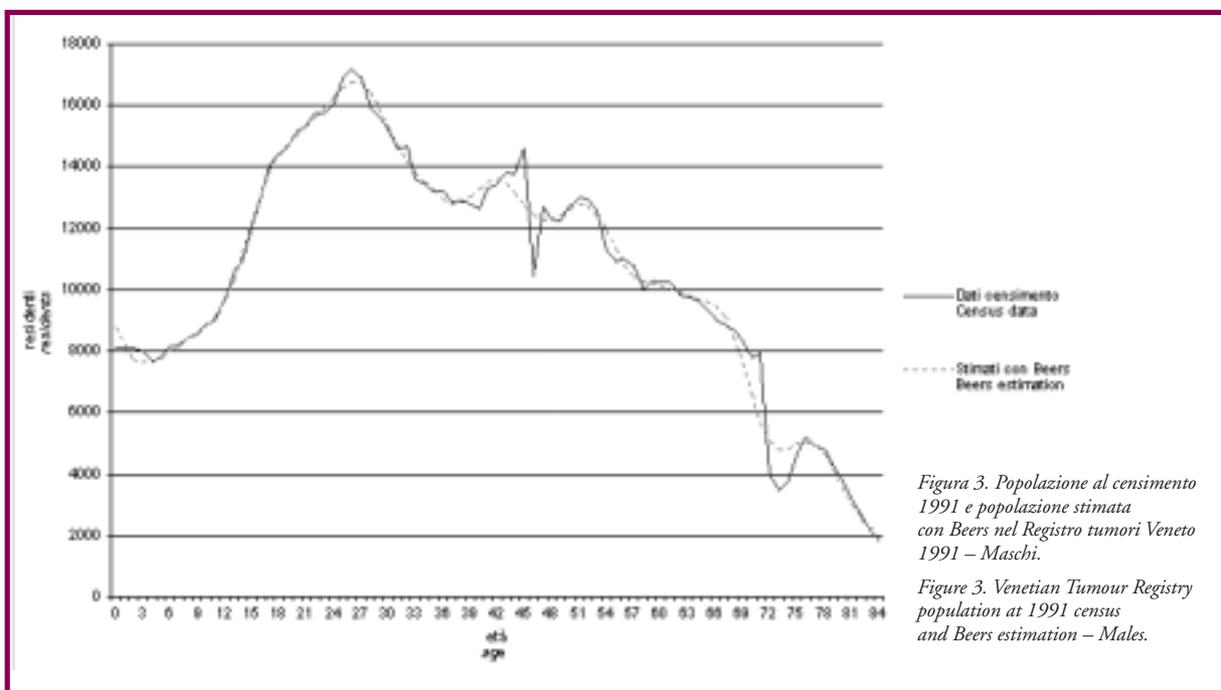


Figura 3. Popolazione al censimento 1991 e popolazione stimata con Beers nel Registro tumori Veneto 1991 – Maschi.

Figure 3. Venetian Tumour Registry population at 1991 census and Beers estimation – Males.

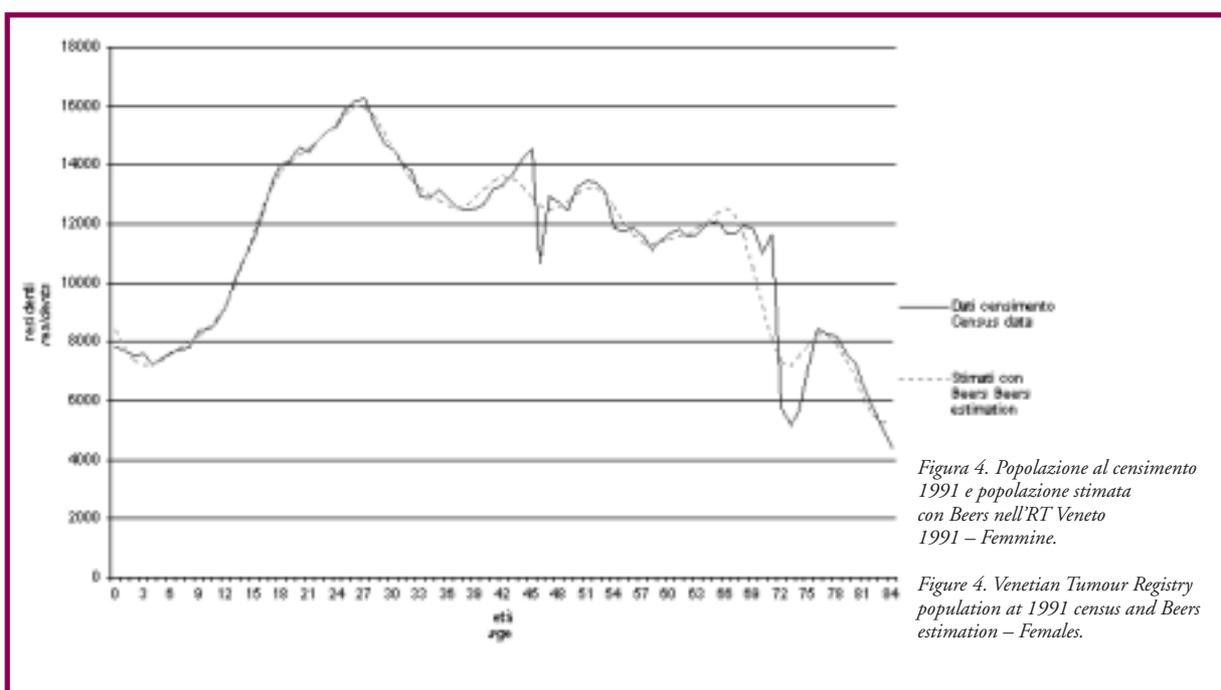


Figura 4. Popolazione al censimento 1991 e popolazione stimata con Beers nell'RT Veneto 1991 – Femmine.

Figure 4. Venetian Tumour Registry population at 1991 census and Beers estimation – Females.

zione e dall'altro al fatto che nella dinamica di popolazione i nati, con le loro fluttuazioni, costituiscono, in assenza di cospicui fenomeni migratori, la principale fonte di rinnovamento della popolazione.

In generale, nelle diverse popolazioni saggiate, le statistiche di bontà di adattamento in tabella 3 mostrano che l'interpolazione finale è buona, tuttavia nelle classi di età più anziane l'a-

dattamento è minore e a volte, come nel caso delle donne ultrasessantacinquenni a Firenze-Prato, Genova e Parma, insufficiente. Tale fenomeno è da attribuire a differenze generate da fluttuazioni confinate a periodi brevi e a un numero limitato di coorti di nascita: anche se il test è particolarmente sensibile a queste fluttuazioni, esse non hanno grossa influenza su tassi aggregati per classi d'età almeno quinquennali.

La possibile distorsione del dato di incidenza o di mortalità a motivo dell'utilizzo di dati di popolazione derivati dall'ultimo censimento disponibile piuttosto che l'uso di stime della popolazione è stata valutata per i dati di incidenza del biennio 1995-96 nel Veneto. I tassi grezzi per tutti i tumori maligni eccetto quelli della pelle non melanocitari, applicando come denominatore la popolazione media 1995-96 piuttosto che la popolazione al censimento 1991, si discostano di non più di un punto percentuale in entrambi i sessi. La differenza tra i due metodi di calcolo dei tassi è invece considerevole per i tassi età specifici, particolarmente per le classi d'età giovani e anziane. Nelle classi giovani il denominatore più basso per i dati del 1995-96 comporta un incremento del valore del tasso non trascurabile: +14% nei maschi e +15% nelle femmine nella classe 0-19 anni; il contrario si verifica nelle classi più anziane in cui la maggiore numerosità della popolazione 1995-96 comporta un decremento nei tassi: -11% nei maschi e -10% nelle femmine nelle età sopra i 65 anni. La diversità nei tassi età specifici si riflette nella standardizzazione per età: utilizzando la popolazione standard mondiale,¹⁰ per gli uomini il tasso standardizzato diminuisce del 9% mentre nelle donne cala del 6%.

Conclusioni

Lo studio sui *trend* di incidenza e mortalità promosso dall'AIRT² ha posto la necessità di costituire un archivio contenente per ogni RT le popolazioni osservate dagli stessi al 31 dicembre di ogni anno, suddivise per sesso e classi annuali d'età.

Questo dettaglio delle informazioni consente agli utilizzatori la massima libertà di elaborazione, permettendo di procedere alle opportune aggregazioni per età e/o periodo secondo la finalità dello studio. Per esempio nelle analisi età-periodo-coorte le classi annuali consentono una maggiore flessibilità nella definizione dell'ampiezza delle celle, specie quando il periodo osservato non è ancora molto esteso, come accade appunto per il *pool* dei dati degli RT italiani.

Sono stati utilizzati tre diversi metodi di interpolazione per la costruzione di un archivio di dati affidabili e dettagliati partendo da dati aggregati per periodo o classi d'età.

Tale archivio informatizzato possiede il seguente tracciato *record*:

- nome registro (8 caratteri);
- anno di calendario (valore numerico);
- classe annuale d'età (valore numerico);
- popolazione residente di sesso maschile al 31 dicembre (valore numerico);
- popolazione residente di sesso femminile al 31 dicembre (valore numerico).

Ogni riga dell'archivio contiene dunque il numero di maschi e il numero di femmine residenti al 31 dicembre di un determinato anno, in un certo RT e per una certa classe annuale d'età (da 0 a 84 anni, con la classe finale aperta di 85 anni e più).

Nelle aree coperte dagli RT si riscontrano le situazioni più di-

verse: l'informazione più frequentemente mancante era quella del dettaglio delle classi annuali d'età, per cui si è ricorso per otto RT al metodo di stima di Beers partendo dalle classi quinquennali. Per tre RT si è utilizzata la stima *pro-rata* per determinare la classe finale d'età di 85 e più anni e per due RT si è ricorso all'interpolazione di Waring, non essendo disponibili i dati di popolazione per l'intero periodo di incidenza coperto.

L'obiettivo di questo lavoro è stato fornire dati di popolazione affidabili da utilizzare essenzialmente come denominatori per l'analisi dei fenomeni, quali l'incidenza o la mortalità per tumore, riferiti alla popolazione residente in un RT. Data la natura di stima e non di osservazione, esso non ha la pretesa di costituire alcuna valenza giuridico-amministrativa.

Le stime sono indispensabili in situazioni come quelle relative alla copertura di aree non istituzionali, diverse cioè da comuni, province o regioni, che nemmeno le più recenti iniziative di monitoraggio dell'ISTAT riescono a soddisfare: alcuni RT coprono una o più ASL (Napoli, Veneto e Romagna), aggregazioni di comuni (Trento) oppure si tratta di ricostruire la popolazione storica di aree di recente definite province (Biella).

Il problema della stima delle informazioni mancanti è stato affrontato anche negli studi di sopravvivenza europei per quanto riguarda i dati di mortalità:

- nello studio Eurocare¹¹ quando erano disponibili solo i tassi di mortalità quinquennali si sono stimati i tassi annuali ipotizzando un incremento costante di mortalità per età all'interno della classe quinquennale;

- nello studio Eurocare-2¹² se mancava completamente l'informazione sulla mortalità per alcuni anni veniva stimata attraverso un metodo di interpolazione lineare indipendente dall'età; nel caso mancasse l'informazione solo per 1 o 2 anni venivano assunti gli stessi tassi di mortalità dell'ultimo anno precedente disponibile.

E' da sottolineare il fatto che i metodi di interpolazione discussi possono essere utilizzati non soltanto per la stima delle popolazioni, ma anche per la stima dei numeratori dei tassi, come casi di malattia o decessi.⁷

Generalmente è sottovalutata l'importanza dell'adeguatezza dei denominatori nei tassi, ossia di avere una corretta misura della popolazione, intesa come numerosità totale e distribuzione per classi d'età, da cui insorgono i casi. Il confronto tra tassi età specifici ottenuti con le popolazioni censuarie e tassi ottenuti con le popolazioni stimate relative allo stesso periodo in cui sono insorti i casi, fornisce una misura dell'influenza che ha la scelta del denominatore. Le differenze maggiori si hanno sulle classi d'età giovani e anziane, per le quali il *trend* di cambiamento nella struttura di popolazione è più importante.

L'archivio informatizzato delle popolazioni sarà reso disponibile sul sito dell'associazione degli RT italiani (<http://www.registri-tumori.it/>) in formato Excel[®] o SAS¹³ secondo il trac-

ciato *record* precedentemente indicato. E' a disposizione inoltre, per chi ne farà richiesta agli autori, un foglio Excel® che permette di inserire una popolazione per classi quinquennali d'età e ottenere la stima delle classi annuali.

Conflitti di interesse: nessuno

Bibliografia

- Buratta V, Marsili M. Il trattamento della mancata risposta in una indagine amministrativa: la rilevazione sulla struttura della popolazione per sesso, anno di nascita e stato civile nei comuni italiani. In: SIEDS-ISTAT, Giornate di studio su «La qualità dell'informazione statistica - Roma, 6-7 aprile 2000».
- Crocetti E, Capocaccia R, Casella C *et al*, eds. Andamenti temporali della patologia oncologica in Italia: i dati dei registri tumori (1986-1997). *Epidemiol Prev* 2004; 28(2) suppl: 1-112.
- Shryock HS, Siegel JS, Stockwell Eg. *The method and materials of demography*. New York, NY: Academic Press, Inc. 1976.
- Johnston J. *Econometrica*. Milano, Franco Angeli Editore, 1993.
- Beers HS. Six-term formulas for routine actuarial interpolation. *The Record of the American Institute of Actuaries* 1945; 34, part 1(69): 59-60.
- Tarone RE, Chu KC. Evaluation of birth cohort patterns in population disease rates. *Am J Epidemiol* 1996; 143(1): 85-91.
- National Center for Health Statistics. *U.S. Decennial life tables for 1989-91*. Center for Disease Control and Prevention. Vol. 1, n. 2.
- Scott DW. *Multivariate density estimation: theory, practice, and visualization*. New York, John Wiley & Sons, 1992.
- Registro Tumori del Veneto. *Incidenza dei Tumori nella Regione Veneto 1995-'96. Cancer Incidence in the Veneto Region 1995-'96*. A cura di: Simonato L, Zambon P, Guzzinati S *et al*. Padova, CLEUP, 1999.
- Segi M. *Cancer Mortality for Selected Sites in 24 countries (1950-57)*. Sendai, Japan, Tohoku University of Medicine, 1960.
- Berrino F, Sant M, Verdecchia A, Capocaccia R, Hakulinen T, Esteve J, eds. *Survival of cancer patients in Europe. The EURO-CARE Study*. Lyon, IARC Scientific Publication, 1995. N.132.
- Berrino F, Capocaccia R, Esteve J *et al*. *Survival of Cancer Patients in Europe: the EURO-CARE-2 Study*. Lyon, IARC Scientific Publication, 1999. N.151.
- SAS Statistical software, v.8.02, SAS Institute, Cary NC, USA



IN BREVE

● Pecore sentinella per studiare la contaminazione da asbesto e fibre asbestiformi

Negli ultimi anni, *Epidemiologia & Prevenzione* ha dedicato spazio all'epidemia di mesoteliomi nel comune di Biancavilla, sulle falde dell'Etna, causata dall'uso - anche per la costruzione di abitazioni - di pietre provenienti dalla locale cava (Montecalvario), contenenti fibre asbestiformi del tipo delle fluoroedeniti. Ricercatori dell'Istituto superiore di sanità e della ASL di Adriano (CT) hanno esaminato i polmoni di 27 pecore con microscopia elettronica a scansione e microanalisi con raggi X e hanno riscontrato la presenza di fibre di fluoroedenite in 8 animali.

La ricerca di fibre inalabili nei polmoni delle pecore sembra quindi essere uno strumento utile per approfondire fenomeni di inquinamento locale. Lo studio conferma precedenti osservazioni in altre aree contaminate con amianto, a Cipro, in Capadocchia e in Corsica (De Nardo P *et al*. *Sci Total Env* 2004; 325: 51-58.)

● Presenza di amianto nell'ambiente generale di Libby, Atlanta

In più occasioni, negli ultimi anni, sono state

espresse preoccupazioni per le possibili conseguenze sulla salute della contaminazione ambientale tra i residenti a Libby, nel Montana (USA), dove da tempo viene estratta vermiculite contaminata con amianto. Sono stati resi noti ora i risultati di uno studio, condotto dai CDC di Atlanta, con un disegno a sezione trasversa, condotto nel 2000-2001 su 7.307 persone che avevano vissuto, lavorato o giocato a Libby per almeno 5 mesi prima del 1990. I partecipanti hanno completato un questionario e a quasi tutti i soggetti di oltre 18 anni è stata effettuata una radiografia del torace. Anormalità pleuriche e interstiziali polmonari sono state rilevate rispettivamente nel 18% e in meno di 1% dei partecipanti. Sono state considerate 29 sorgenti di esposizione, nell'ambiente di lavoro, nelle abitazioni e in altre circostanze.

La prevalenza di anormalità pleuriche aumentava in modo correlato al numero di sorgenti di esposizione, da 6,7% in soggetti che non segnalavano esposizione a 34,6% in coloro che indicavano 12 o più sorgenti di esposizione (Peipins LA *et al Environ Health Perspect* 2003; 111: 1753-59).