

L'uso dei modelli per la stima del trend complessivo di incidenza

The use of models for estimating overall incidence trend

Riccardo Capocaccia,¹ Emanuele Crocetti²

¹ Istituto superiore di sanità, Laboratorio di epidemiologia, Roma

² Registro tumori toscano. Unità operativa di epidemiologia clinica e descrittiva. Centro per lo studio e la prevenzione oncologica, Istituto scientifico della Regione Toscana, Firenze

Corrispondenza: Riccardo Capocaccia, Istituto superiore di sanità, Laboratorio di epidemiologia, viale Regina Elena 299, 00161 Roma

Riassunto

Nell'ambito dello studio sui trend tumorali dell'Associazione Italiana Registri Tumori sono state analizzate varie possibilità per stimare un trend che possa includere Registri con periodi di attività diversi nell'ambito di uno stesso arco temporale. Lo studio è stato condotto sui nove Registri del pool AIRT per il periodo 1986-1997. I dati per il periodo 1986-1993 del Registro di Ragusa non sono stati utilizzati per simulare l'inclusione di un Registro con osservazioni parziali. Sono stati utilizzati quattro diversi approcci: 1) *Golden standard* (GS) calcolato come media pesata tra i tassi standardizzati (popolazione Europea) osservati nel pool AIRT; 2) *Trend grezzo* (TG) calcolato nel pool AIRT con dati parziali per Ragusa; 3) *Trend modellizzato* (TM) stimato applicando un modello log-lineare con età, registro e periodo come variabili indipendenti al set di dati parziali; *Trend corretto* (TC) calcolato sulla base dei dati completi degli otto registri, dei dati osservati a Ragusa nel quadriennio 1994-1997 e di quelli di Ragusa stimati dal modello (3) per il 1986-1993.

Le stime dei trend, espresse come variazione media annua percentuale dei tassi standardizzati (EAPC), sono state confrontate rispetto al GS. I risultati di questa analisi preliminare sconsigliano l'uso del trend grezzo ed indicano il trend corretto quale metodo più affidabile per la stima degli andamenti.

Introduzione

I registri tumori italiani rilevano l'incidenza in una popolazione di oltre tredici milioni di persone, pari al 23,4% della popolazione italiana. È noto che tale popolazione non costituisce un campione rappresentativo della popolazione nazionale, essendo prevalentemente concentrata nelle regioni del Centro nord. Si tratta comunque di una popolazione demograficamente ben definita, su cui vengono rilevati anche altri indicatori epidemiologici, quali la sopravvivenza, la mortalità, ecc. Ai fini di una migliore comprensione delle tendenze recenti delle malattie neoplastiche è quindi utile poter analizzare i trend di incidenza nel complesso della popolazione coperta dai registri tumori italiani. La stima di tale andamento complessivo deve tenere conto del fatto che l'attività di registrazione non copre lo stesso periodo in tutte le aree geografiche interessate. I dati di incidenza inclusi della

Abstract

Several approaches to evaluate an overall trend including Registries with different length of activity with a defined period of time have been evaluated in the cancer trend study of the Italian Network of Cancer Registries (AIRT). The study used the nine Registries of the pool AIRT in the period 1986-1997. Data for 1986-1993 from the Registry of Ragusa have been excluded to simulate the inclusion of a Registry with partial observation.

Four different approaches have been used: *Golden standard* (GS) as a weighted mean of the observed standardised (European population) rates in the whole pool AIRT; 2) *Raw trend* (TG) computed with partial data of Ragusa; 3) *Model trend* (TM) estimated by means of a log-linear model including age, registry and period as independent variables and applied on the partial data set; *Adjusted trend* (TC) computed on complete data for eight Registries, observed 1994-1997 data of Ragusa and estimated data for Ragusa for 1986-1993 computed by (3) model.

The trend estimates, as estimated annual percent change of the standardised rates (EAPC) have been compared with GS. The results of this preliminary analysis do not recommend the use of raw trend and indicate adjusted trend as the most reliable method for trend estimate.

Introduction

Italian cancer registries are collecting the incidence data in a population of over thirteen millions inhabitants, accounting for 23.4% of the total Italian population. It is known that such a sample is not representative of the national population, being mostly concentrated in northern and central regions. However it is a demographically well defined population, from which other epidemiological indicators are collected, such as survival, mortality, etc. In order to better understand recent cancer trends, it is useful to analyse incidence trends in the population covered by cancer registries. The estimate of such overall trend must take into account that registration does not cover the same calendar period in single involved geographical areas. Incidence data in the Italian Network of Cancer Registries (AIRT) database cover a period from 1976 to 1996. The final year of registration is 1997 for most registries, 1998 for four, and 1996 for two registries. A solid estimate of incidence time trends needs to

Site	Males				Females			
	GS	TG	TM	TC	GS	TG	TM	TC
Oesophagus	-4.0	-8.2	-1.4	-4.6	-1.0	-5.7	-0.2	-0.4
Stomach	-0.7	-4.2	-1.1	-0.7	0.2	-3.9	-0.9	0.7
Ileum	5.5	2.4	6.5	6.4	7.0	2.2	2.3	4.0
Colon	4.1	0.7	4.1	4.3	3.2	-0.1	2.3	3.2
Rectum	3.0	0.7	2.0	2.3	1.1	-1.9	1.5	0.9
Liver	2.3	0.5	4.2	2.2	3.3	1.2	4.1	4.3
Biliary tracts	0.2	2.3	3.2	1.8	1.5	2.4	1.4	2.0
Pancreas	4.0	1.9	3.2	4.0	3.4	-0.2	3.2	3.5
Larynx	-1.8	-6.6	-2.1	-2.6	1.1	-3.5	-0.6	1.1
Lung	0.3	-3.0	0.5	0.3	2.1	-1.9	2.4	2.7
Bone	-0.3	0.0	-1.2	0.7	-2.6	-2.2	-1.6	-1.5
Melanoma	7.8	4.8	6.9	8.4	5.6	1.0	5.9	5.4
Breast					2.6	-0.7	2.5	2.4
Cervix uteri					1.6	4.5	-1.7	1.1
Corpus uteri					2.4	-1.4	2.3	2.8
Uterus n.o.s.					0.1	0.5	-0.9	-2.8
Ovary					5.4	3.8	2.6	4.6
Prostate	10.2	6.5	8.1	9.8				
Testis	3.5	-1.8	1.7	2.3				
Kidney	3.4	-1.6	4.1	3.0	2.2	-1.8	4.4	3.4
Bladder	5.5	1.6	3.3	3.6	5.3	0.1	4.1	4.5
CNS	2.7	0.6	1.4	1.5	2.3	-1.2	2.0	1.1
Thyroid	9.1	3.5	3.8	5.2	9.8	5.8	4.7	7.3
Hodgkin	-2.2	-3.4	-3.1	-1.6	-2.9	-2.5	-2.1	-1.0
Non-Hodgkin	3.8	0.6	5.0	4.2	4.4	-0.2	4.8	4.7
Myeloma	6.5	4.7	5.0	6.8	5.5	2.1	4.5	5.5
Leukemias	4.9	3.2	2.4	4.4	1.4	0.0	1.9	3.5
All sites	2.9	0.6	2.4	2.7	2.5	0.0	2.2	2.7

Tabella 1. Variazione annua percentuale del tasso standardizzato (EAPC) di incidenza nel periodo 1994-1997 osservata (GS = golden standard) e stimata con tre diversi metodi: trend grezzo (TG), trend modellizzato (TM) e trend corretto (TC) per sesso e sede tumorale. In grassetto sono indicati i valori che più si avvicinano a quelli di riferimento (GS).

Table 1. Annual percent variation of the incidence standardized rate (EAPC) during 1994-1997: observed (GS = golden standard) and estimated with three different methods: Raw Trend (TG), Model Trend (TM) and Adjusted Trend (TC) by sex and cancer site. The values most similar to the reference (GS) are reported in bold.

banca dati dell'Associazione Italiana Registri Tumori (AIRT) coprono un periodo il cui inizio va dal 1976 al 1996. L'anno finale di registrazione è per quasi tutti i registri il 1997, per quattro registri il 1998, e per due il 1996. Una stima consistente del *trend* temporale di incidenza richiede ovviamente la considerazione di una popolazione omogenea all'interno dell'arco temporale considerato. La struttura dei dati disponibile richiede perciò un compromesso tra il grado di rappresentatività geografica, restringendo il periodo analizzato, e la lunghezza del periodo, escludendo però i registri con periodo di rilevazione più breve. L'uso di modelli statistici permette di risolvere in parte questa alternativa utilizzando, al costo di opportune assunzioni supplementari, tutta l'informazione disponibile. In questo lavoro saranno esplorati e discussi diversi metodi potenzialmente utilizzabili per ottenere una stima del *trend* complessivo di incidenza nelle aree dei registri. A tale scopo sarà innanzitutto predisposta una situazione sperimentale semplificata, costruita sui dati effettivamente osservati dei registri, ed adatta per meglio evidenziare le caratteristiche dei diversi metodi utilizzati. Questi saranno applicati ad un insieme di 28 sedi tumorali, distinte

consider a homogeneous population within the considered period. The structure of available data forced to a compromise between the degree of geographic representativity, by reducing the analysed period, and the length of such a period, by excluding those registries with a shorter registration period. The use of statistical models allows to solve, at least partially, such a compromise, by some supplementary assumptions, using all the available information. In this paper we will explore and discuss several methods which could be used to obtain an estimate of the overall incidence trend in the Registries area. With this purpose we will first define a simplified experimental scenario, built from real data observed by the Registries, which will better allow to outline the characteristics of the different methods used. These will be applied on a pool of 28 different cancer sites, stratified by gender. Results will be then analysed and evaluated according to the different trends shown by single cancer types in the studied period.

Material

The study has been limited to the incidence period 1986-1997, the same period used for the time trend analysis

per sesso. I risultati saranno quindi analizzati e valutati anche in relazione ai diversi andamenti che i tumori considerati hanno presentato nel periodo in esame.

Materiali

Lo studio è stato limitato al periodo di incidenza 1986-1997, lo stesso periodo di base utilizzato per l'analisi temporale presentata in questa monografia. Sono stati quindi selezionati i nove registri con dati di incidenza sostanzialmente completi per tutta la durata del periodo. Il *trend* di incidenza calcolato sull'insieme di questi registri (*pool* AIRT) è stato assunto come *golden standard* con cui confrontare le diverse stime. Queste sono state applicate ad un set di dati modificato cancellando artificialmente i dati di incidenza di uno dei registri per tutto il periodo 1986-1993. È stato scelto a tale scopo il registro di Ragusa, che presenta i valori di incidenza più differenti dalla media generale (per lo più inferiori), e che quindi contribuisce al *pool* modificato con soli dati relativi al quadriennio 1994-1997. Infine, allo scopo di mettere meglio in evidenza le differenze tra i *trend* ottenuti con i diversi metodi di stima ed il *golden standard* di riferimento, il peso di Ragusa all'interno del *pool* dei nove registri è stato aumentato come se la sua popolazione fosse il 25% (anziché il 3%) di quella complessiva. La situazione sperimentale sopra descritta è stata predisposta per 23 sedi tumorali negli uomini e 26 sedi tumorali nelle donne (Tabella 1). Per semplicità di analisi e di rappresentazione, gli anni di calendario sono stati raggruppati in tre quadrienni: 1986-1989, 1990-1993, 1994-1997.

Metodi

Quattro diversi *trend* sono stati stimati per ciascuna sede e per ciascun sesso, utilizzando come indicatore in tutti i casi il tasso standardizzato sulla popolazione europea.

Il *Golden standard* (GS) è stato calcolato dalla media pesata tra i tassi standardizzati osservati nel complesso degli otto registri (peso 0,75) e quelli osservati, non troncati, a Ragusa (peso 0,25), per ciascun quadriennio.

Il *Trend* Grezzo (TG) è stato calcolato dal set di dati modificato aggregando tutti i dati disponibili, senza pertanto considerare la diversa rappresentatività geografica nei diversi periodi analizzati (otto registri nel 1986-1993, nove registri nel 1994-1997).

Il *Trend* Modellizzato (TM) è stato stimato applicando un modello log-lineare con età, registro e periodo come variabili indipendenti al set di dati modificato:

$$\log(r_{ikt}) = a_i + b_k + c_t + \text{ }_{kjt} \quad (1)$$

dove r_{ikt} è il tasso di incidenza specifico per la classe di età quinquennale i , l'area k ($k=1,2,n$) e l'anno t ; a , b e c sono rispettivamente i parametri relativi alla classe di età, al registro ed al periodo, ed il tempo t è espresso dal tempo medio

presented in this monograph. We have then selected the nine Registries with substantially complete incidence data for the whole period (pool AIRT). The incidence trend calculated on the pool AIRT has been assumed as a golden standard to compare different estimates. Estimates have been applied to a dataset modified by erasing artificially incidence data from one registry for the whole 1986-1993 period. The registry of Ragusa has been chosen for such a purpose, as it showed incidence data with the greatest variation (mostly lower) from the average, and thus contributes to the modified pool only with data from the 1994-1997 period. Finally, to achieve a better evidence on the difference between trends obtained with different estimate methods and the reference golden standard, the weight of Ragusa registry within the pool has been increased as if its population did account for 25%, rather than the real 3%, of the overall population. The described experimental scenario has been defined for 23 cancer sites in men and 26 cancer sites in women (Table 1). To simplify the analysis and the presentation of results, calendar years have been grouped in three four-year periods: 1986-1989, 1990-1993, 1994-1997.

Methods

Four different trends have been estimated for each cancer site and by gender, using the European standardised rate as a reference in all cases.

The Golden standard (GS) has been calculated according to the weighted average of standardised rates observed in the pool of eight registries (weight 0.75) and of those observed, not truncated, at Ragusa (weight 0.25) for each four-year period.

The Raw Trend (TG) has been calculated from the dataset, modified by aggregating all available data, taking no care of the different geographical representativity (eight registries during 1986-1993, nine registries during 1994-1997) in the considered periods.

The Model Trend (TM) has been estimated by applying to the modified dataset a log-linear model, with age, registry and period as independent variables:

$$\log(r_{ikt}) = a_i + b_k + c_t + \text{ }_{kjt} \quad (1)$$

where r_{ikt} is the specific incidence rate for the 5-year age category, k is the area ($k=1,2,n$), and t is the year; a , b and c are parameters related to age categories, registry and period, respectively, whereas time t is the average time of each 4-year period.

Finally, the Adjusted Trend (TC) has been calculated on the basis of complete data from eight registries, on data from Ragusa observed during 1994-1997, and on those estimated for Ragusa by the model (1) for the two previous 4-year periods. In other words, the model here is used only as a method of estimating lacking data for the registry of Ragusa.

di ciascun quadriennio. Infine, il *Trend Corretto* (TC) è stato calcolato sulla base dei dati completi degli otto registri, dei dati di Ragusa osservati nel quadriennio 1994-1997 e di quelli di Ragusa stimati dal modello (1) nei due quadrienni precedenti. In altre parole, il modello è qui utilizzato solo come metodo di imputazione dei dati mancanti per il registro di Ragusa. La variazione percentuale media annua (o *Estimated Annual Percent Change* – EAPC) del tasso di incidenza standardizzato per il quadriennio 1994-1997 è stata utilizzata come indicatore per il confronto delle stime ottenute con i diversi metodi sperimentati.

Risultati

La Tabella 1 riporta il valore delle EAPC ottenuto per ciascuna combinazione sede-sesso e con ciascuno dei quattro metodi. Come sopra ricordato, i valori forniti dal metodo GS sono da considerare come i valori di riferimento su cui valutare le stime ottenute con gli altri metodi. I valori in grassetto sono quelli che, nell'ambito dei tre metodi utilizzati, più si avvicinano ai valori di riferimento. Su un totale di 49 combinazioni sede-sesso, il metodo del TC fornisce la migliore stima in 37 casi, quello del TM in 9 casi, ed il metodo del TG in soli 3 casi. È anche importante valutare in quanti casi ciascuno dei metodi abbia fornito una stima molto diversa dal *golden standard* sia in termini relativi che in termini assoluti. Sono state quindi definite come non accettabili delle stime di EAPC inferiori o superiori del 50% a quello reale (in termini di differenze relative) e delle stime di EAPC maggiori o minori di quattro punti percentuali (in termini di differenze assolute) rispetto al valore reale. Si è infine valutato quante volte ciascun metodo forniva un *trend* di segno opposto rispetto a quello reale. La Tabella 2 dà un quadro riassuntivo delle valutazioni effettuate con i diversi criteri. Tutti i criteri utilizzati sconsigliano l'uso del *trend* grezzo ed indicano il *trend* corretto quale metodo più affidabile per la stima degli andamenti.

Discussione

Questo articolo presenta alcune valutazioni circa diversi metodi di stima di un *trend* temporale nel caso in cui dati relativi a intervalli temporali diversi siano disponibili nelle diverse aree geografiche considerate. I risultati di questo lavoro debbono essere considerati assolutamente preliminari: analisi più rigorose sono necessarie per poter dare un significato più generale alle indicazioni sopra espresse. Il modello (1) è stato utilizzato per la stima del *trend* complessivo con il metodo TM e, limitatamente all'imputazione dei dati mancanti, anche con il metodo TC. Esso si basa sull'ipotesi di un andamento comune per tutte le aree, sia per quelle con dati completi che per quelle con dati mancanti. La verifica di tale assunzione, condotta con un appropriato test statistico, dovrebbe essere un preliminare necessario per l'applicazione del modello nell'ambito dei metodi descritti. Un'analisi più

	TG	TM	TC
Best estimate as compared to golden standard	3	9	37
Estimate not acceptable in relative terms	40	19	9
Estimate not acceptable in absolute terms	15	3	0
Trend inversion	13	2	0

Tabella 2. Valutazione complessiva dell'aderenza delle stime ottenute con i diversi metodi rispetto alla variazione annuale osservata per il periodo 1994-1997. Numero di casi, su un totale di 49 combinazioni sede-sesso, in cui veniva soddisfatto ciascun criterio di valutazione. *Trend grezzo* (TG), *trend modellizzato* (TM) e *trend corretto* (TC).

Table 2. Overall evaluation of the difference of estimates obtained with different methods as compared to annual variation observed during 1994-1997. Number of cases, on a total of 49 site/gender combinations, in which each evaluation criterion was fulfilled. Raw Trend (TG), Model Trend (TM) and Adjusted Trend (TC).

The Estimated Annual Percent Change (EAPC) of standardised incidence during 1994-1997 has been used as a reference to compare estimates obtained with different experimental methods.

Results

Table 1 shows EAPC values obtained for each site-gender combination and with each of the four methods. As previously mentioned, values obtained with the GS method must be considered as reference values, on which estimates obtained with other methods should be evaluated. Values reported in bold are the most similar to reference values within those calculated with the other three methods. Out of 49 site/gender combinations, the TC method provides the best estimate in 37 cases, the TM method in 9, and the TG only in three cases. It is also important to evaluate in how many cases each method provided a much different estimate as compared to the golden standard, either in relative or absolute terms. EAPC estimates lower or higher (50% in terms of relative difference or 4% in terms of absolute difference) as compared to the real rate were considered not acceptable. Finally we have evaluated how often each method estimated an opposite trend as compared to the real one. Table 2 summarises estimates performed with different criteria. All used criteria do not recommend the use of the raw trend and indicate adjusted one as the most reliable method for trend estimate.

Discussion

This paper discusses different methods used to estimate a time trend when data recorded during different calendar periods are available from different areas. The results of this paper must be considered very preliminary: more complete analyses are needed to validate the previously discussed indications. The model (1) has been used for the estimate of the overall trend with the TM method and also with the TC method to estimate lacking data. The method is based on the hypothesis of a common trend for all areas, those with complete and those with incomplete ones. Such assumption should be

approfondita sulle sedi in cui il metodo del *trend* corretto (TC) ha dato risultati inaccurati mostra in tutti i casi una disomogeneità del *trend* osservato a Ragusa rispetto a quello degli altri registri. Le altre assunzioni di indipendenza implicite nel modello, quella tra età e registro e quella tra età e periodo, sono forse meno influenti sulla validità del risultato finale. Tuttavia, la sensibilità delle stime a tutte queste ipotesi di indipendenza deve essere rigorosamente valutata. Il modello utilizzato assume un andamento lineare del logaritmo del tasso di incidenza. Nelle applicazioni reali dovranno essere utilizzati modelli più generali, di tipo non lineare o *jointpoint*. Un andamento semplicemente lineare dovrà essere considerato solo sulla base di test statistici appropriati. Infine, nel confronto tra i risultati dei diversi metodi non si è tenuto conto della variabilità statistica. Una valutazione rigorosa dovrà essere basata, oltre che sulle stime puntuali, anche sui relativi intervalli di confidenza.

In conclusione, l'analisi condotta nel presente lavoro permette di proporre una procedura di stima del *trend* temporale in grado di fornire delle stime affidabili, con il massimo utilizzo dei dati osservazionali e con il minimo impatto del contributo fornito dalle stime modellistiche. Tale procedura può essere schematizzata nei seguenti passi:

1. Stima di un modello non lineare del tasso specifico di incidenza con età, area e periodo come variabili indipendenti. A tale scopo, le aree con identica copertura temporale possono essere aggregate.
2. Valutazione statistica dell'omogeneità del trend tra le aree geografiche.
3. Nel caso di rigetto dell'omogeneità, stima di un modello separato per area geografica.
4. Valutazione della linearità del trend complessivo o di ciascun trend area-specifico.
5. Utilizzo dei modelli stimati per le aree con dati mancanti per l'imputazione di questi ultimi.
6. Aggregazione finale di tutte le aree e calcolo del trend complessivo.

Questa procedura dovrà essere comunque vagliata da un successivo e più rigoroso studio di validazione.

verified with a proper statistical test, as a necessary preliminary condition to apply the model within the described methods. A more detailed analysis of those sites for which the TC method gave inaccurate results, shows in all cases that the trend observed for Ragusa is not homogenous as compared to other registries. The other assumptions of independence intrinsic to the model, between age and registry and between age and period, may have a lower influence on the validity of the final result. Nevertheless, the sensitivity of estimates to all these hypotheses of independence must be thoroughly evaluated. With the used model the incidence rate (on a logarithmic scale) has a linear trend. More general models, not linear or of jointpoint type, will be necessarily used in real applications. A simple linear trend will have to be considered only on the basis of proper statistical tests. Finally, when comparing the results obtained with different methods, we did not take into account the statistical variability. A detailed evaluation will have to be based, apart from punctual estimates, also on the relative confidence interval.

In conclusion, the analysis performed in the present study allows to suggest a procedure to estimate time trends which seems to provide reliable estimates, with the most complete use of observational data, and with a minimal impact of the contribution of model estimates. Such a procedure may be summarised as follows:

1. *Estimate of a non-linear model of specific incidence rate, with age, area and period as independent variables. For this purpose, areas with the same time coverage may be aggregated.*
2. *Statistical evaluation of the homogeneity of the trend between geographical areas.*
3. *In case of no homogeneity, estimate of a separate model for each geographical area.*
4. *Evaluation of the linearity of the overall trend or of each area-specific trend.*
5. *Use of models to estimate rates for the areas with lacking data.*
6. *Final aggregation of all areas and estimate of the overall trend. This procedure will have to be checked with a future, more complete validation study.*