

I tumori in Italia - Rapporto 2007

**Materiali, metodi
e indicatori di qualità**

Italian cancer figures - Report 2007

**Materials, methods
and quality indicators**

Materiali e metodi

La banca dati dei registri tumori italiani

La banca dati dei registri tumori italiani raccoglie attualmente i dati di 22 registri generali di popolazione e 4 registri specializzati (tumori infantili, mesotelioma e colon-retto), comprendo così una vasta area del territorio nazionale, con una maggior prevalenza nel Nord e Centro Italia (Figura 1). Prima di essere abilitati a fornire dati, i registri attraversano un processo d'accreditamento che tiene conto anche degli aspetti istituzionali e strutturali necessari a sostenere nel tempo un lavoro di qualità adeguata. In seguito, in occasione di ogni aggiornamento, i dati sono ulteriormente controllati secondo una serie di procedure concordate. I protocolli per l'accreditamento e il controllo di qualità possono essere consultati sul sito web dell'Associazione italiana registri tumori (AIRTUM) che gestisce direttamente la banca dati attraverso i propri organi societari e i gruppi tecnici (www.registri-tumori.it). I registri sono nati in anni differenti, e quindi la copertura nel tempo è cambiata. Nella Figura 2 sono delineati per ogni registro gli anni di rilevazione dell'incidenza, gli anni successivi d'osservazione dello stato in vita dei pazienti (follow-up) e la selezione operata per quest'analisi.

Criteri di selezione

Le analisi di sopravvivenza qui presentate riguardano i casi di primi tumori primitivi maligni diagnosticati dal 1985 al 2002. Sono stati quindi inclusi unicamente i casi con codice di comportamento /3 e /6 dell'ICD-O (3rd ed).¹ Sono stati esclusi i carcinomi della cute (codice ICD-10 C44).² Sono stati inoltre esclusi i secondi tumori primitivi, con l'eccezione di quelli insorti dopo un carcinoma della cute. Come è d'uso per lo studio della sopravvivenza su base di popolazione, ovvero per i casi registrati secondo i criteri internazionali che regolano i registri tumori, sono stati inoltre esclusi casi identificati mediante il solo certificato di morte (DCO), oppure mediante autopsia. Al contrario, i casi identificati molto vicini alla data di morte, ovvero con sopravvivenze inferiori a 30 giorni, sono stati inclusi nell'analisi.

Le sedi da analizzare sono state scelte con il duplice obiettivo di consentire confronti con studi precedenti e proporre qualche approfondimento di alcune sedi più specifiche oppure tipi istologici particolari. Alcune sedi sono state analizzate e presentate sia separatamente, sia raggruppate, in modo da facilitare il confronto con altri studi che abbiano utilizzato tali raggruppamenti.

Per la scelta del periodo di diagnosi e delle aree di registro da analizzare sono stati considerati due diversi criteri di selezione.

■ **Dati 1995-1999.** Il fuoco dell'analisi è sui casi diagnosticati dall'inizio del 1995 alla fine del 1999, cui contribuiscono, con

Materials and methods

The Italian cancer registries database

The Italian Cancer Registries Database currently collects data from 21 General Population Registries and 4 Specialised Registries (Childhood, Mesothelioma and Colon-Rectum), thus covering a large area of the national territory, with a better coverage in Northern and Central Italy (Figure 1). First, the Registries undergo an accreditation process, which also takes into account the institutional and structural aspects necessary to sustain an adequate data gathering quality over time; then, every time an update is done, data are checked again according to a series of agreed procedures. Accreditation protocols and quality control procedures can be found on the website of the Italian Network of Cancer Registries (AIRTUM) which directly manages the Database, through its organisational bodies and technical groups (www.registri-tumori.it). Different Registries were started in different years and thus their coverage has changed over time. Figure 2 shows for each Registry the year when incidence was recorded, the follow-up year and the selection made for the data analysis presented in this paper.

Selection criteria

The survival analysis presented in this paper is based on first pri-

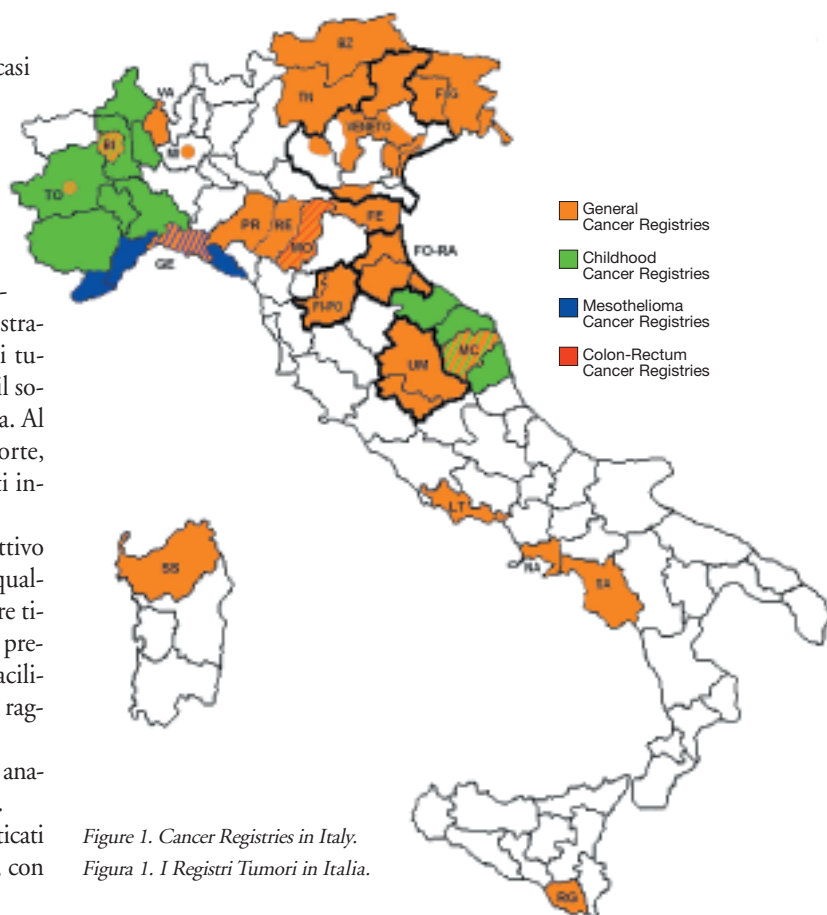


Figure 1. Cancer Registries in Italy.
Figura 1. I Registri Tumori in Italia.

Registry	1976	1977	1978	1979	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Torino										■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Biella																														
Genova																														
Varese																														
Milano																														
Veneto																														
Trento																														
Alto Adige																														
Friuli Venezia Giulia																														
Parma																														
Reggio Emilia																														
Modena																														
Ferrara																														
Romagna																														
Firenze-Prato																														
Umbria																														
Macerata																														
Napoli																														
Salerno																														
Sassari																														
Ragusa																														

Figure 2. Available years of incidence (■) and follow-up (■) by Registry. The orange line indicates the registries chosen for cohort analysis of survival. The blue line indicates the registries and the years chosen for trend analysis.

Figura 2. Anni d'incidenza (■) ed anni di follow-up (■) disponibili per Registro. La linea arancio indica i registri scelti per l'analisi di sopravvivenza di coorte. La linea blu indica i registri e gli anni scelti per l'analisi del trend.

l'eccezione di Milano (vedi commento pag. 22) tutti i registri inclusi nella banca dati AIRTUM. La data comune di chiusura del follow-up è stata posta al 31 dicembre 2003. Ciò consente di avere un sufficiente periodo di osservazione per calcolare la sopravvivenza a 5 anni dalla diagnosi. I soli casi diagnosticati nel 1999 hanno infatti un follow-up potenziale di 4 anni. Studi precedenti (Eurocare3) hanno mostrato che la conseguente distorsione è trascurabile o inesistente.

■ **Trend temporali 1985-2002.** Si è inoltre voluto analizzare anche di quanto la sopravvivenza è migliorata a partire dal 1985, da quando cioè un numero sufficiente di registri era in grado di fornire dati di qualità: per questo abbiamo inserito un'apposita sezione in cui sono presentati i trend temporali di sopravvivenza dei casi diagnosticati dal 1985 al 2002. Per questa analisi sono stati selezionati i dati dei 7 registri attivi nel corso di tutto il periodo considerato (Torino, Genova, Veneto, Parma, Romagna, Firenze-Prato, Ragusa). Anche qui, la data di chiusura del follow-up è il 31 dicembre 2003.

Metodi di analisi

Gli indicatori di sopravvivenza per le coorti di pazienti diagnosticati fino al 1999 sono stati calcolati sulla base delle infor-

mary malignant tumours diagnosed from 1985 to 2002. Only cases with a code of behaviour /3 and /6 of the ICD-O (3rd ed.)² were included. Skin carcinomas (code ICD-10¹ C44) were excluded. Moreover, second primary tumours were excluded, with the exception of those that arose after a skin cancer. Cases identified only through death certificate (DCO) or autopsy were excluded, as by the established practice in the study of survival in a population-based setting, which entails the recording of cases according to the international rules that regulate cancer Registries. Instead, cases identified very close to the date of death, with a survival period below 30 days, were included in the data analysis.

The sites to be analysed were chosen with the double objective of allowing comparison with previous studies and proposing a more in-depth study of some specific sites or particular histological types. Some sites were analysed and presented both separately and pooled, in order to facilitate comparison with other studies that use the same pooling criteria.

Two different selection criteria have been considered for choosing the diagnosis period and the Registry area to be analysed.

■ **Data 1995-1999.** This analysis focuses on the cases diagnosed from the beginning of 1995 to the end of 1999, with the contribution of all Registries included in the AIRTUM database with

mazioni sullo stato in vita degli stessi pazienti raccolte fino al 31 dicembre 2003. Le sopravvivenze osservate sono state calcolate secondo il metodo attuariale, le sopravvivenze relative secondo il metodo proposto da Hakulinen,³ utilizzando il programma SEER*Stat (release 6.2.4) fornito dal Programma statunitense di sorveglianza e controllo dei tumori dell'Istituto nazionale tumori americano (The Surveillance Research Program of the Division of Cancer Control and Population Science of the National Cancer Institute). Maggiori dettagli sono disponibili presso il sito: <http://seer.cancer.gov/seerstat>. L'errore standard della sopravvivenza relativa cumulativa è calcolato secondo la tradizionale formula:

$$ES_x^{CR} = CR_x \sqrt{\sum_{j=1}^x \frac{D_i}{L_i^+ (L_i^+ - D_i)}}$$

dove CR_x è la sopravvivenza relativa cumulativa sino all'intervallo di tempo x ; D_i è il numero di casi deceduti nell'intervallo di tempo i e L_i^+ è il numero di persone vive all'inizio dell'intervallo x , corrette per il numero di casi persi e censiti nell'intervallo i .

L'errore standard è stato poi utilizzato per calcolare i limiti di confidenza al 95% come: indicatore ± 1.96 *errore standard. Nel caso i limiti superassero il dominio di 0 ed 1, sono stati troncati a tali valori.

La sopravvivenza relativa standardizzata per età è stata calcolata suddividendo i casi in ampie classi d'età per dare maggior stabilità statistica all'indicatore. Le classi sono: 15-44, 45-54, 55-64, 65-74, 75+. Lo standard usato è quello proposto da Corazziari et al.⁴ che è composto da 4 tipologie di standard, a seconda del tipo di tumore, o meglio della distribuzione per età dei casi incidenti (Tabella 1). Lo standard di tipo 1 è associato ai tumori che si verificano prevalentemente in età anziana, quindi è quello utilizzato nella maggior parte delle sedi. Gli standard 2 e 3 corrispondono rispettivamente a tumori con incidenza approssimativamente equidistribuita per età (melanoma, cervice uterina, encefalo), e a quelli con prevalenza di casi in età giovanile (testicolo, malattia di Hodgkin, leucemia linfatica acuta). Lo standard 4 è utilizzato esclusivamente per il tumore della prostata. L'errore standard è calcolato come:

$$ES_x^{CRx} = \sqrt{\sum_{k=1}^{K-5} w_k^2 * ES_{x,k}^{CR}}$$

dove $ES_{x,k}^{CR}$ è l'errore standard della sopravvivenza relativa all'intervallo x , per la classe d'età k , e w_k è il peso per la k -esima classe d'età.

L'analisi degli andamenti temporali in Italia include anche una

the exception of the Milan Registry (see comment on page 22). The common date of follow-up ending was 31st December 2003. This allows this study to have a sufficiently long follow-up period to calculate survival at 5 years from diagnosis. Only cases diagnosed in 1999 show a potential 4-year follow-up. Previous analyses [Eurocare3] show that consequent distortion is negligible or non-existent.

■ **Time trend 1985-2002.** We also decided to analyse how much survival has improved since 1985, that is from when a sufficient number of registries were able to provide quality data: thus we included a section presenting the survival trends for cases diagnosed from 1985 to 2002. This analysis includes data from 7 Registries (Torino, Genova, Veneto, Parma, Romagna, Firenze-Prato, Ragusa) which were active throughout the considered period. Even in this case, the follow-up ended on 31st December 2003.

Analysis methods

Survival indicators for cohorts of patients diagnosed up to 1999 have been calculated on the basis of information on vital status of patients till 31st December 2003. Observed survival was calculated according to the actuarial method, and relative survival according to the method proposed by Hakulinen³, utilising the SEER*Stat 6.2.3 programme, provided by the Surveillance Research Program of the Division of Cancer Control and Population Science of the National Cancer Institute, USA (more details are available at: <http://seer.cancer.gov/seerstat>). The standard error of the cumulative relative survival is calculated according to the traditional formula:

$$ES_x^{CR} = CR_x \sqrt{\sum_{j=1}^x \frac{D_i}{L_i^+ (L_i^+ - D_i)}}$$

where CR_x is the cumulative relative survival up to the time interval x ; D_i is the number of deceased cases in the time interval i and L_i^+ is the number of people alive at the beginning of interval x , adjusted for the number of cases lost and recorded in the interval i . The standard error was then used to calculate the 95% confidence limits, as follows: indicator ± 1.96 *standard error. If the limits fall outside the domain of 0 and 1, they were truncated at those values.

To give it more statistical stability, relative survival standardised by age was calculated by dividing the cases into large age groups, namely: 15-44, 45-54, 55-64, 65-74, 75+. The standard used is that proposed by Corazziari et al.⁴ which is composed of 4 types of standards, depending on the type of tumour, or more precisely on the distribution of its incidence by age (table 1). Type 1 standard is associated to tumours that set in mainly during old age, thus it is used in the majority of sites. Standards 2 and 3 correspond respectively to tumours with the same incidence for each age group by age group (melanoma, cervix uteri, brain, ...), and

Standard	Age groups	Weights	Cancer sites
1	15-44, 45-54, 55-64, 65-74, 75+	7, 12, 23, 29, 29	all (except those included in standards 2,3,4)
2	15-44, 45-54, 55-64, 65-74, 75+	28, 17, 21, 20, 14	nasopharynx, soft tissues, melanoma, cervix uteri, brain, thyroid, bone.
3	15-44, 45-54, 55-64, 65-74, 75+	60, 10, 10, 10, 10	testis, Hodgkin disease, acute lymphatic leukaemia
4	15-54, 55-64, 65-74, 75-84, 85+	19, 23, 29, 23, 6	prostate

Table 1. Standard population used to calculate standardised relative survival by sex and age.

Tabella 1. Popolazione Standard utilizzata per il calcolo della sopravvivenza relativa standardizzata per età.

stima della sopravvivenza nel triennio 2000-02. Tale stima è ottenuta con la metodologia di Brenner,⁵ implementata nel software SEER*Stat, e permette di valutare in anticipo, rispetto ai metodi di calcolo convenzionali, i risultati prognostici raggiunti negli anni recenti. Le stime di periodo si basano infatti sull'esperienza di vita dei pazienti osservata e raggruppata in funzione del tempo trascorso dalla diagnosi, per gli anni più recenti a disposizione: nel nostro caso per il periodo 2000-2003. Mentre la stima convenzionale, o di coorte, è necessariamente differita rispetto al periodo di osservazione dell'incidenza, la stima di periodo permette di utilizzare appieno i dati più recenti anche nello studio della sopravvivenza a lungo termine, cioè a 10 anni o 15 anni dalla diagnosi. E' bene comunque tener presente che le stime di periodo non si sostituiscono all'analisi per coorte di diagnosi, ma ne rappresentano una anticipazione o proiezione. I risultati sono presentati secondo tavole specifiche per sede e/o tipo di tumore. Abbiamo scelto per esse alcuni indicatori che, a nostro avviso, meglio sintetizzavano quanto emerge da questa analisi, e in particolare:

■ **Tabella dei dati pool:** in questa tabella sono presentate per sesso e classi d'età, il numero di casi all'inizio del follow-up, le sopravvivenze osservate e relative per l'insieme dei registri generali attualmente attivi in Italia e che hanno fornito dati utilizzabili per questa analisi (Tabella 2). Gli intervalli di tempo presentati sono quelli che, si ritiene, possano al meglio segnalare eventuali differenze di sopravvivenza, ovvero a 1 anno per i tumori a breve sopravvivenza, a 3 e a 5 anni per i tumori a media e lunga sopravvivenza, anche se, per questi ultimi, ormai il taglio a 5 anni può sembrare ancora troppo breve per apprezza-

those prevailing at a younger age (testicle, Hodgkin, acute lymphatic leukaemia). Standard 4 is used exclusively for prostate cancer. The standard error is calculated as:

$$ES_x^{CR} = \sqrt{\sum_{k=1}^{K-5} w_k^2 * ES_{x,k}^{CR}}$$

where $ES_{x,k}^{CR}$ is the standard error for relative survival calculated at the interval x , for the age group k , and w_k is the weight for the k -nth age group.

The analysis of time trends in Italy also include an estimate of survival between 2000-2002. This estimate was obtained through the Brenner method⁵ implemented in the SEER*Stat software. In relation to the conventional calculation methods, this allows for an earlier evaluation of the prognostic results reached in recent years. Period estimates are based on the available life experience of the patients diagnosed during the most recent years, while the conventional cohort estimate is necessarily 5 years behind the time of observation of incidence. A more timeous estimate is particularly useful in the study of long-term survivals, which is the proportion of patients alive after 10 or more years from diagnosis. It must be born in mind that period estimates cannot replace diagnosis by cohort analysis, but provide an anticipation or a projection. Results are presented in specific tables either by site and/or by type of tumour. We chose some indicators that in our opinion best summarise the results of this analysis, in particular:

Registry	Registries included in the 1995-1999 cohort analysis	Area	Registries included in the trend analysis
Torino	yes	North	yes
Biella	yes	North	
Genova	yes	North	yes
Varese	yes	North	
Milano	yes	North	
Veneto	yes	North	yes
Trento	yes	North	
Alto Adige	yes	North	
Friuli Venezia Giulia	yes	North	
Parma	yes	Centre-North	yes
Reggio Emilia	yes	Centre-North	
Modena	yes	Centre-North	
Ferrara	yes	Centre-North	
Romagna	yes	Centre-North	yes
Firenze-Prato	yes	Centre-North	yes
Umbria	yes	Centre-North	
Macerata	yes	Centre-North	
Napoli	yes	South	
Salerno	yes	South	
Sassari	yes	South	
Ragusa	yes	South	yes

Note: South includes Sicily and Sardinia

Table 2. Registries included in the analyses.

Tabella 2. Registri inclusi nelle analisi.

re eventuali differenze. Nel caso delle sopravvivenze a 5 anni sono anche forniti gli intervalli di confidenza al 95% calcolati come descritto sopra.

■ **Grafici per registro:** gli indicatori di sopravvivenza relativa a 5 anni, standardizzata per età, sono rappresentati per i diversi registri italiani per mezzo di un grafico a barre, ordinati per livello di sopravvivenza. I colori utilizzati per le barre consentono di distinguere i registri operanti nelle aree del Nord, Centro-Nord e Sud Italia. Una linea posta a livello del valore dell'indicatore indica i suoi limiti di confidenza al 95%. Nelle sedi frequenti, i grafici vengono presentati separatamente per sesso. Per quelle meno frequenti, e nel caso in cui non vi fosse una significativa differenza di sopravvivenza tra uomini e donne nell'insieme dei dati, viene presentato un unico grafico per i due sessi. Allo scopo di presentare statistiche sufficientemente solide, nei grafici non vengono rappresentati i registri per cui sono stati osservati meno di 20 casi incidenti per la sede (e il sesso) considerati. Singoli registri sono stati anche necessariamente esclusi dal confronto quando l'assenza di casi in una delle classi di età rendeva impossibile il calcolo della sopravvivenza standardizzata. Infine, i registri con numerosità compresa fra 20 e 50 casi all'inizio del follow-up, sono stati indicati con un asterisco.

In questo grafico manca il registro di Milano, di recente istituzione, che raccoglie i casi solo a partire dal 1999. Al 31 dicembre 2003 non si raggiungono infatti i canonici 5 anni di osservazione, l'intervallo di tempo scelto per i confronti degli indicatori di sopravvivenza. Tuttavia in una tabella separata vengono forniti gli indicatori di sopravvivenza a 1 e 3 anni per poter effettuare il confronto al meglio, senza il rischio di distorsioni dovute alla differente struttura dei dati.

■ **Grafico per area:** la curva di sopravvivenza relativa standardizzata fino agli 5 anni di osservazione è presentata per le aree del Nord, Centro-Nord e Sud Italia (isole comprese). Essa consente di apprezzarne non solo il livello a intervalli di tempo definito, ma anche l'andamento, la cui variazione può fornire importanti suggerimenti sul fenomeno in studio.

■ **Grafico sui trend di sopravvivenza:** si confrontano le sopravvivenze relative standardizzate a 5, 10 e 15 anni dal 1985 al 1999 divisi in periodi triennali. Naturalmente, all'aumentare del tempo d'osservazione, diminuiscono i periodi in cui è possibile avere un'osservazione completa. Inoltre per gli anni 2000-02 è anche fornita la stima di periodo che consente, come ricordato, di avere un'idea dei futuri miglioramenti della sopravvivenza per i casi diagnosticati più recentemente, di cui tuttavia non si dispone ancora di un tempo d'osservazione sufficientemente lungo.

References - Bibliografia

1. Fritz A, Percy C, Jack A, Shanmugaratnam K, Sobin L, Parkin DM, Whelan S, (eds). *International classification of disease for Oncology. Third Edition*, World Health Organization, Geneva 2000.
2. World Health Organization. *International statistical classification of diseases and related health problems. Tenth Revision*, Geneva 1992.
3. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; 38: 993-42.
4. Corazziari I, Quinn M, Capocaccia R. Standard cancer patients population for age standardising survival ratios. *Eu J Cancer* 2004; 40: 2307-16.
5. Brenner H, Hakulinen T. Advanced detection of time trends in long-term patient survival: experience from 50 years of cancer registration in Finland. *Am J Epidemiol* 2002; 156(6):566-77.

■ **Pool table:** this table shows the number of cases from the beginning of the follow-up, observed and relative survivals by gender, by age group for the total of the general Registries presently active in Italy and that have provided usable data for this analysis (table 2). The time intervals shown are those that we believe are most suitable to show possible differences in survival, these are 1 year for short-term survival tumours, 3 and 5 years for medium- and long-term survival, although, for the latter, the 5-year cut-off point may still appear too short to capture possible differences. In the case of survival over 5 years, the 95% confidence intervals are also provided, calculated as described above.

■ **Graph by registry:** 5-year survival indicators, standardised by age, are shown for the different Italian Registries through a bar chart, ordered by survival level. The bar colours allow to identify the Registries in the North, Centre-North, and South of Italy. A line at the value level of the indicator shows its 95% confidence limits. For frequent sites the graphs are presented separately by sex. For the less frequent ones, and when there was no significant difference in survival for men and women, only one graph is presented for both sexes jointly. In order to present sufficiently robust statistics, if there were less than 20 incident cases for a site (and gender) these registries are not shown. Single registries for which there was a lack of cases in one of the age groups that made survival calculation impossible were also excluded. Finally, registries with a record of between 20 and 50 cases at the beginning of the follow-up are marked with an asterisk.

This graph does not include the Milan Registry, which since it has been recently established only has recorded cases since 1999, and thus lacks the required 5-year follow-up, the time interval chosen for comparison in this study. However, indicators for 1 and 3 years are presented separately, to better make a comparison without the risk of distortion due to different methods.

■ **Graph by area:** The relative standardised survival curve up to 5 years of follow-up is shown for the North, Centre-North, and South (including Sicily, and Sardinia) of Italy. This curve not only shows survival at specific time intervals, but also the survival trend, whose variations may provide important insight about the phenomenon studied.

■ **Survival trend graph:** This graph compares relative standardised survival at 5, 10, and 15 years from 1985 to 1999, divided into 3-year intervals. Naturally, as the follow-up time increases, the period for which a complete follow-up is available decreases. An estimate for the period 2000-2002 that gives an idea about the future survival improvement for recently diagnosed cases is also provided, although the follow-up time is not yet sufficiently long.