

## La banca dati AIRTUM

### Introduzione

L'Associazione italiana registri tumori raccoglie i dati da Registri tumori di popolazione, sia generali sia specializzati, e, dopo averne verificato qualità e completezza, li utilizza per attività collaborative descrittive e di ricerca sull'epidemiologia oncologica nel nostro Paese.<sup>1-6</sup>

In Italia l'attività di registrazione si è sviluppata a partire dagli anni Settanta con una graduale e costante crescita delle esperienze e della quota di popolazione interessata dalla rilevazione dei tumori.

Attualmente l'Associazione raccoglie 31 Registri, dei quali 26 generali e 5 specializzati (per età o per tipologia tumorale).

La concentrazione dei Registri è maggiore al Nord (specialmente Nord-Est) rispetto al Centro e soprattutto all'Italia meridionale e insulare. D'altra parte è proprio nel Sud Italia che in anni recenti si è avuto il maggior sviluppo di nuovi Registri, che ha permesso di documentare più dettagliatamente e con miglior rappresentatività la patologia oncologica anche in quest'area.

Nella Tabella 1 è presentata la quota di popolazione osservata dall'Airtum per singola regione e per macroaree geografiche (Nord, Centro, Sud). La diversa presenza e la varia dimensione dei Registri determinano una variazione nella copertura regionale che va dallo 0% per diverse regioni (Calabria, Puglia, Basilicata, Abruzzo, Molise e Valle d'Aosta), fino al 100% (per esempio Umbria, Friuli Venezia Giulia e le province autonome di Trento e Bolzano). Nel meridione, nonostante vi siano ancora numerose regioni senza Registri, si è avuta la maggior crescita di nuove esperienze: la copertura, che per il Sud Italia nel 1989 era dell'1,4% (per la presenza del solo Registro di Ragusa), è attualmente passata al 15,7%. Negli ultimi 20 anni la popolazione registrata nel Centro Italia è più che raddoppiata passando la 10,7% al 25,7% così come al Nord, dove si è passati dal 21,9% al 48,0%.

In diverse regioni lo sviluppo dei Registri è avvenuto proprio in questi 20 anni con crescite notevoli sia al Nord e al Centro, sia nel Sud Italia, per esempio in Sardegna e in Campania, dove la copertura della popolazione è aumentata di 28,5 punti percentuali, e in Sicilia, dove è cresciuta di 16,9 punti. Inoltre, nuove realtà nel Sud Italia e nell'Italia insulare hanno già iniziato la fase di accreditamento dei dati presso l'Airtum facendo prevedere un prossimo ulteriore incremento della popolazione interessata e una estensione della registrazione anche a regioni attualmente sprovviste di Registri.

Complessivamente i Registri dell'AIRTUM tengono sotto controllo i dati oncologici di quasi 19.000.000 di italiani, pari al 32,2% della popolazione residente totale. Come accennato, queste percentuali sono del 48,0% nel Nord, del 25,7% nel Centro e del 15,7% nel Sud Italia (Tabella 1).

## The AIRTUM database

### Introduction

*The Italian Network of Cancer Registries (AIRTUM) collects data from both general and specialized population-based cancer registries. AIRTUM verifies data quality and completeness and uses data for collaborative studies on cancer epidemiology in Italy.<sup>1-6</sup>*

*Cancer registration in Italy started in the 70s and has gradually extended, involving a growing proportion of the Italian resident population. At this time, the AIRTUM database collects data from 26 general and 5 specialized (for age or cancer type) cancer registries.*

*The density of registries is greater in the north of Italy, especially in the north-east, in comparison with the centre and the south. However, cancer registration has recently undergone significant development in the south of Italy, providing a wider dataset for this area.*

*Table 1 presents the proportion of the resident population involved in cancer registry activity, for each region and for three geographic areas.*

*Regional coverage varies from 0% in several southern regions (Calabria, Puglia, Basilicata, Abruzzo, Molise), as well as Val d'Aosta in the north, up to 100% (e.g., Umbria, Friuli Venezia Giulia, Trento and Bolzano). Nevertheless, there has been an increase in cancer registration in the south of Italy. In 1989, population coverage was 1.4% (there was only one registry, in Ragusa), whereas now coverage is 15.7%. Over the last twenty years, population coverage has more than doubled in central Italy, from 10.7% to 25.7%, and in northern Italy, where the percentage has grown from 21.9% to 48.0%.*

*Over the last twenty years, there has been an increase in the number of registries in several regions, in the north and centre but also in the south, e.g., in Sardinia and in Campania registry coverage has increased by 28.5%, in Sicily by 16.9%. Moreover, new registries in the south of Italy have already asked to be included in AIRTUM and their data are currently being reviewed. Therefore, in the near future the Italian resident population will be better represented, including in regions where at the moment there are no registries.*

*Overall, the AIRTUM registries cover about 19,000,000 subjects, 32.2% of the entire Italian resident population, 48.0% in the north, 25.7% in the centre, and 15.7% in the south of Italy, Table 1.*

*Besides general registries covering the entire population and all types of cancer, AIRTUM also includes specialized registries for cancers in childhood (0-14 years) and teen-age (15-19 years), in the regions of Piedmont and Marche, which cover about 700,000 children and 150,000 teenagers. Moreover, AIRTUM includes registries that collect data for specific cancer types: the colorectal cancer registry of Modena, the Liguria registry of mesothelioma, and the registry of female breast cancer of Palermo. Specialized cancer registries overall cover more than 2,300,000 subjects, therefore the overall population observed by AIRTUM is 36% of the Italian population.*

Oltre ai Registri generali che interessano tutta la popolazione e tutti i tipi di tumore, all'AIRTUM partecipano anche Registri specializzati per i tumori infantili (anni 0-14) e adolescenziali (anni 15-19) nella Regione Piemonte nelle Marche, che interessano complessivamente circa 700.000 bambini e 150.000 ragazzi. Fanno parte dell'AIRTUM anche tre registri specializzati in singole patologie tumorali: il Registro dei tumori coloretali di Modena, quello dei mesoteliomi della Regione Liguria e il Registro del tumore della mammella femminile di Palermo. Nel complesso l'attività dei Registri specializzati interessa oltre 2.300.000 abitanti, portando la copertura totale della rete AIRTUM al 36% della popolazione italiana.

### La banca dati

Nel 2005 l'AIRTUM si è dotata di una banca dati che centralizza le informazioni prodotte dai singoli Registri. La banca dati, che rientra tra gli obiettivi statutari dell'Associazione, ha avuto il merito formalizzare la tipologia di dati raccolti in un formato record flessibile, che permette il continuo aggiornamento sia dei casi già noti (integrando nuove variabili clinico-prognostiche, o informazioni su eventuali nuovi tumori insorti nello stesso soggetto), sia dello stato in vita. Ogni Registro mantiene la proprietà dei propri dati e il loro utilizzo per studi collaborativi necessita di specifica autorizzazione.

Nel sito dell'Associazione sono disponibili sia il protocollo operativo con la descrizione dettagliata dei criteri di inclusione, della tipologia e delle modalità dei controlli dei formati degli archivi, sia le regole per l'accesso alla banca dati (<http://www.registri-tumori.it/cms/?q=manuali#>).

### Criteri di inclusione

La banca dati è composta da tre archivi diversi:

- il primo raccoglie i dati di incidenza prodotti dai Registri,
- il secondo contiene i dati di mortalità (fonte Istat o ReN-Cam, Registro nominativo delle cause di morte di Usl),
- il terzo contiene i dati riguardanti le popolazioni residenti nelle aree di registrazione.

La banca dati di incidenza raccoglie tutti i tumori maligni incidenti osservati dai Registri, ma anche tumori non maligni (*in situ*, benigni del sistema nervoso centrale, tumori a comportamento incerto) per i Registri che effettuano questa raccolta.

Indipendentemente dalla classificazione con cui i casi incidenti sono codificati da ogni Registro, questi devono essere inviati alla banca dati in accordo con la Classificazione internazionale delle malattie 3<sup>a</sup> edizione (ICD-O 3),<sup>7</sup> oltre che con la classificazione originale, se diversa da questa.

I tumori multipli sono definiti in accordo con la più recente classificazione proposta dalla International Association of Cancer Registries (IACR), dalla International Agency for Research on Cancer (IARC) e dall'European Network of Cancer Registries (ENCR) (<http://www.enrcr.com.fr/>).

AREA	Observed population in 2009 no.	Regional coverage in 2009 %	Regional coverage in 1989 %	Change from 1989 to 2009 %
Piemonte	1,088,710	25.1	23.3	1.9
Valle D'Aosta	0	0.0	0.0	0.0
Lombardia	3,304,823	35.2	8.9	26.2
Trentino Alto Adige	974,613	100.0	0.0	100.0
Veneto	2,294,685	48.8	33.3	15.6
Friuli Venezia Giulia	1,204,718	100.0	0.0	100.0
Liguria	875,732	55.0	42.0	13.0
Emilia Romagna	2,962,001	71.3	40.8	30.6
<b>NORTH</b>	<b>12,705,282</b>	<b>48.0</b>	<b>21.9</b>	<b>26.1</b>
Toscana	1,207,581	33.6	33.1	0.4
Umbria	858,938	100.0	0.0	100.0
Marche	304,443	20.0	0.0	20.0
Lazio	519,850	9.9	0.0	9.9
<b>CENTRE</b>	<b>2,890,812</b>	<b>25.7</b>	<b>10.7</b>	<b>15.0</b>
Abruzzo	0	0.0	0.0	0.0
Molise	0	0.0	0.0	0.0
Campania	1,647,217	28.5	0.0	28.5
Puglia	0	0.0	0.0	0.0
Basilicata	0	0.0	0.0	0.0
Calabria	0	0.0	0.0	0.0
Sicilia	1,133,666	22.6	5.8	16.9
Sardegna	469,870	28.5	0.0	28.5
<b>SOUTH</b>	<b>3,250,753</b>	<b>15.7</b>	<b>1.4</b>	<b>14.3</b>
<b>ITALY</b>	<b>18,846,847</b>	<b>32.2</b>	<b>12.3</b>	<b>19.9</b>

Table 1. AIRTUM database. Population observed by the Italian Network of Cancer Registries (general and specialized), by region and area and percentage change from 1989 to 2009.

### The database

AIRTUM has had a centralized database since 2005, when a unique, flexible database, which is also one of AIRTUM's main formal aims, was created. In this database, as regards incidence cases, each case is identified by a unique identification number which allows future updates of clinical and personal data.

Each registry's Director remains the owner of his/her data; data are sent to the AIRTUM database and used according to specific rules. On the AIRTUM website there is a detailed description of the rules regarding inclusion criteria, checks, and formats, as well as instructions on how to access the database and use the data <http://www.registri-tumori.it/cms/?q=manuali#>

### Inclusion criteria

The AIRTUM database includes three archives, one for incident cancer cases, one for cancer deaths (based on the National Institute of Statistics or on Local Mortality Registries - ReN-Cam) and a third for resident populations.

With respect to incidence, the database collects all cancers gathered by each registry, invasive and not invasive (*in situ*, benign of the central nervous system, of uncertain behaviour). Incident cases have to be sent to the database coded with the original clas-

Table 2. AIRTUM database. All observed cases (except for non-melanoma skin cancer), 1998-2005. Some quality indicators, by sex, year of diagnosis and geographic area. Percentage of microscopically verified (MV) cases (histology and cytology), mortality incidence ratio (M/I) and percentage of cases known from the death certificate only (DCO).

Males	MV (%)			M/I			DCO (%)		
	North	Centre	South	North	Centre	South	North	Centre	South
1998	85	81	78	53	56	56	1	1	2
1999	85	81	75	53	57	58	1	1	3
2000	85	82	76	51	54	55	1	1	3
2001	86	83	77	51	54	56	1	1	3
2002	86	82	78	50	54	55	1	1	2
2003	87	85	80	49	52	54	1	1	2
2004	88	88	80	49	51	52	1	0	2
2005	88	89	81	49	51	54	0	1	2
Females	MV (%)			M/I			DCO (%)		
	North	Centre	South	North	Centre	South	North	Centre	South
1998	86	81	84	47	50	48	2	2	2
1999	86	82	81	46	48	44	1	2	3
2000	86	82	82	47	49	47	1	1	3
2001	87	84	82	47	47	45	1	1	3
2002	87	84	83	46	45	46	1	1	3
2003	87	84	83	47	47	45	1	1	3
2004	88	87	85	46	46	48	1	1	3
2005	89	89	87	46	47	48	1	0	2

### Controlli

I dati di incidenza destinati alla banca dati sono sottoposti a due ordini di controlli di qualità, previsti dal protocollo operativo.

Il primo, precedente l'invio dei dati, viene effettuato dal singolo Registro, che è tenuto a sottoporli ai *check* previsti dal programma DepEdit prodotto dalla International Agency for Research on Cancer ([www.iarc.com.fr/](http://www.iarc.com.fr/)), verificando che non vi siano segnalazioni di errore e certificando che i *warning* segnalati sono stati verificati e confermati.

Il secondo a livello della banca dati che prevede di fare nuovamente i *check* previsti dal DepEdit oltre alle verifiche di un *software* prodotto dall'Associazione, CheckRT, che controlla la corrispondenza fra i valori presenti e un valore atteso sulla base di un *gold standard* rappresentato dalla casistica della banca dati, eventualmente corretto per area geografica. Vengono effettuati complessivamente 1.360 controlli su 17 gruppi di tipologie di possibili errori. Il software è gratuitamente disponibile nel sito AIRTUM (<http://www.registri-tumori.it/cms/?q=software>)

Inoltre, a livello centralizzato sono effettuati controlli di congruenza e completezza delle variabili del tracciato.

Infine, a livello centrale sono anche effettuati controlli particolari previsti per la partecipazione a specifici studi e non inclusi nelle precedenti fasi di verifica.

Prima di ogni pubblicazione i dati di ogni Registro sono sottoposti ad alcune delle verifiche di qualità previste per la procedura di accreditamento all'AIRTUM.

### La banca dati utilizzata nel Rapporto AIRTUM 2009

Nel presente studio sono stati inclusi i dati relativi ai tumori maligni infiltranti, presenti in banca dati a maggio 2009 e relativi al periodo 1998-2005. Sono stati inclusi i registri che avevano dati per tutto il periodo di osservazione o almeno per 6 degli 8 anni analizzati, in questo ultimo caso i dati mancanti sono stati stimati secondo il metodo descritto nella sezione *Materiali e Metodi* (pp. 151).

Nel complesso sono stati analizzati 818.017 casi incidenti e 342.444 decessi oncologici.

La sezione *Dati AIRTUM 1998-2005* presenta il contributo

*sification adopted by the registry at the time of case inclusion and also according to the 3rd version of the International Classification of Diseases for Oncology (ICD-O 3).<sup>7</sup>*

*Multiple primaries are coded according to the newest classification proposed by the International Association of Cancer Registries (IACR), the International Agency for Research on Cancer (IARC) and the European Network of Cancer Registries (ENCR) (<http://www.enrcr.com.fr/>).*

### Controls

*Each registry, before sending its own data to the database, must check data completeness and quality. Controls are based on the 'DepEdit' software produced by the IARC ([www.iarc.com.fr/](http://www.iarc.com.fr/)). Errors must be corrected and warnings checked and corrected or confirmed.*

*'DepEdit' checks and other specific checks are also performed centrally, as well as a number of data consistency and completeness checks. In particular, a specific software provided by AIRTUM itself, CheckRT, is used to compare single registry data with a gold standard. It performs 1,360 checks of 17 different types of errors. This software is available on the AIRTUM website (<http://www.registri-tumori.it/cms/?q=software>)*

*Finally, other checks required for participation in specific studies are also performed centrally.*

*Before data of a registry are included in a publication of AIRTUM, their quality is verified according to the main rules for registry inclusion in AIRTUM.*

### Database used for the 2009 AIRTUM Report

*In this report we analysed all malignant invasive cases included in the database as of May 2009, incident during 1998-2005. Registries with data for the entire period, 1998-2005, or for at least 6 years were included and lacking data were es-*



in termini di casi osservati e di decessi, per sede tumorale e per sesso, di ciascun registro incluso nell'analisi di questa monografia (pp. 156).

### Indicatori di qualità

Nella Tabella 2 sono indicati per sesso, anno di incidenza e per macroaree geografiche alcuni indicatori di qualità dei dati osservati dai Registri nel periodo 1998-2005.

Nello specifico per il totale dei tumori (esclusi gli epiteliumi della cute) è presentata:

- la percentuale di casi con conferma microscopica (MV) sia istologica sia citologica;
- il rapporto mortalità/incidenza (M/I);
- la percentuale di casi noti dal solo certificato di morte (DCO).

La presenza di una conferma microscopica assicura che la patologia identificata dal Registro sia proprio un caso di tumore. In linea teorica questa conferma dovrebbe tendere al 100%. In realtà non tutti i casi tumorali fanno anche accertamenti diagnostici invasivi o trattamenti chirurgici ai quali seguono gli esami anatomopatologici, basti pensare a quadri che arrivano alla diagnosi in fase avanzata, soprattutto in pazienti anziani. In questi casi la conferma del patologo non aggiungerebbe informazioni utili né per il trattamento né per la prognosi. Inoltre una quota di pazienti può aver avuto la conferma microscopica lontano dall'area di residenza, non rendendo queste informazioni disponibili al Registro; quest'ultimo aspetto è particolarmente rilevante nel meridione, dove si osserva una importante migrazione sanitaria.<sup>8</sup>

In media nel 2005 la percentuale di casi con conferma microscopica nell'archivio AIRTUM è risultata pari a 87 per gli uomini e a 88 per le donne, con valori inferiori nel Sud Italia tra gli uomini (81) e più omogenei tra le donne. La percentuale è in crescita nelle tre aree.

Il rapporto mortalità/incidenza è un indicatore di completezza che si basa sull'assunto che i dati di mortalità siano completi e che il rapporto con l'incidenza sia stabile. Nel 2005 nel database AIRTUM questo rapporto è risultato pari a 50 tra gli uomini e 46 tra le donne, il rapporto è migliorato (si è ridotto) nel tempo nelle tre aree. Rimangono ancora delle differenze fra Nord, Centro e Sud, soprattutto fra gli uomini, indice presumibilmente di una miglior completezza dei Registri del Centro-Nord ma anche di una peggior sopravvivenza media nel meridione.<sup>2</sup>

Infine, la quota dei casi noti al Registro dal solo certificato di morte (DCO) informa sulla tempestività del Registro nell'intercettare la storia clinica prima del decesso. I valori osservati nel 2005, in media 1% per gli uomini e 1% per le donne, sono nel complesso buoni, anche se migliori nel Centro-Nord. Questo indicatore viene gradualmente sostituito dal numero di casi che giungono all'osservazione del Registro come DCO, che prendono il nome di DCI (Death Certificate Initiated). L'arrivo di un DCI attiva un processo di revisione della docu-

timata as described in the Material and Methods section. Overall, 818,017 incident cases and 342,444 cancer deaths were analysed.

The section Incidence and mortality data: AIRTUM cases shows the number of cases and deaths for each registry, sex and cancer site (pp. 156).

### Quality indicators

Table 2 presents some quality indicators for the period 1998-2005 by sex, year of incidence, and geographic area.

In particular, for all incident cancers (except non-melanoma skin cancers) the table presents the following:

- the percentage of microscopically verified (MV) cases (histology and cytology);
- the mortality/incidence ratio;
- the percentage of cases known from the death certificate only (DCO)

Microscopic verification confirms that a cancer is correctly defined and collected. In principle, it should be 100%. However, not all cancer cases do have aggressive diagnosis or treatment and microscopic verification, especially those which are diagnosed at a late stage and/or in elderly patients. In these cases, microscopic verification does not improve either diagnosis (already evident) or prognosis. Moreover, some patients may have performed microscopic verification far from the place of residence. For these patients microscopic verification exists but it is not available for the registry; this can be often the case in the south of Italy where migration towards hospitals of other northern regions is significant.<sup>8</sup> On average in 2005, MV in the AIRTUM database was 87% for men and 88% for women, with lower values in the south for men (81). MV has improved in all three geographic areas over time.

Mortality / Incidence ratio is an index of completeness based on the assumption that mortality data are complete and the M/I rate is stable. M/I in AIRTUM in 2005 was 50 for men and 46 for women; the rate improved (decreased) among women over time. There are still some differences between north, centre, and south, especially among men. This presumably means that registries in the north and centre are more complete than those in the south of Italy, but it may also be due to the worse cancer survival rate in the south of Italy (AIRTUM WG 2007).

The third quality indicator is the proportion of cases known from the death certificate only (DCO). This index is related to the timeliness of the registry in collecting data on the clinical history of each case before the death.

The mean percentage of DCOs was 1% for men and women in 2005, an overall good result, though the data are better in northern and central Italy than in the south. Instead of DCOs, it would be more useful to evaluate cases recorded for the first time by the death certificate, or death certificate initiated (DCI). A DCI prompts the registry to do a trace-back search for further information; only if no further information can be found is the case recorded as DCO. If, instead, further information (from clinical

mentazione clinica disponibile, di ricerca di nuova documentazione e di contatti con il medico curante che spesso portano nuove informazioni che definiscono il caso in modo diverso da DCO. Quindi in realtà più dei DCO è interessante conoscere il numero dei DCI, ma non tutti i Registri inviano questa informazione alla banca dati.

Nel complesso questi dati indicano una buona qualità globale in termini di completezza e di validità dei casi raccolti. In media appare qualche problema in più per i Registri del meridione, ma non tale da pregiudicare la confrontabilità fra aree.

*reports, general practitioners, etc.) is found, the DCI becomes a regular case. Therefore, the proportion of DCIs is more relevant than that of DCOs to evaluate registry completeness and timeliness. Unfortunately not all the AIRTUM Registries have provided this information to the database.*

*Overall, quality indicators show a general good quality and completeness of data. On average, registries in the south of Italy have slightly worse indicators than the others. However, the average quality of data allows for reliable comparisons among geographic areas.*

### References - Bibliografia

1. Airt Working Group. I tumori in Italia – Rapporto 2006. Incidenza, mortalità e stime. *Epidemiol Prev* 2006; 30(1) suppl 2: 1-148.
2. Airtum Working Group. I tumori in Italia – Rapporto 2007. Sopravvivenza. *Epidemiol Prev* 2007; 31(1) suppl 1: 1-96.
3. Airtum Working Group. I tumori in Italia – Rapporto 2008. I tumori infantili: incidenza, sopravvivenza andamenti temporali. *Epidemiol Prev* 2008; 32(1) suppl 1:1-112.
4. Airtum Working Group. I tumori in Italia – Documento 2009. I nuovi dati di incidenza e mortalità. Periodo 2003-2005. (<http://www.registri-tumori.it/cms/?q=Doc2009>)
5. De Angelis R, Francisci S, Baili P et al. Database on cancer survival in Europe: data standardisation, quality control and methods of statistical analysis. *Eur J Cancer* 2009; Apr;45(6): 909-30.
6. Dal Maso L, Lise M, Zambon P et al. Incidence of primary liver cancer in Italy between 1988 and 2002: an age-period-cohort analysis. *Eur J Cancer* 2008 Jan; 44(2): 285-92.
7. Fritz A, Percy C, Jack A et al. Classificazione internazionale delle malattie per l'oncologia. Terza edizione, versione italiana a cura di Giacomini A, Ferretti S. ([http://www.registri-tumori.it/cms/files/ICDO\\_2005.pdf](http://www.registri-tumori.it/cms/files/ICDO_2005.pdf))
8. Dardanoni G Pollita Addario S. Dinamiche dei ricoveri per tumore della mammella in Sicilia e stima dell'incidenza delle procedure chirurgiche. In: Il tumore della mammella in regione Sicilia. A cura di Airtum, Registri tumori della Sicilia, OER\_Sicilia. *Epidemiol Prev* 2009; 33(1-2) suppl 1: 27-32.