



Capitolo 2

Materiali e metodi

Materials and methods

OBIETTIVI

La Monografia AIRTUM 2014 si pone l'obiettivo di presentare stime aggiornate e proiezioni del numero di persone in vita dopo una diagnosi di tumore (la prevalenza completa). La prevalenza completa dei tumori in una determinata popolazione è definita come la proporzione o il numero di persone a cui è stato diagnosticato un tumore nel corso della vita, e che sono vive a una certa data di riferimento. È una misura che dipende dagli andamenti di incidenza e di sopravvivenza per tumore e dai tassi di mortalità generale nella popolazione considerata. L'invecchiamento della popolazione produce un aumento dei tassi di incidenza per tumore e, quindi, ha un effetto anche sulla prevalenza. La prevalenza dei tumori, dunque, è un indicatore della richiesta di cure sanitarie dei pazienti oncologici e include anche coloro che possono essere considerati guariti e necessitano di limitate risorse aggiuntive da parte del sistema sanitario.

Nella presente Monografia si fornisce un ampio spettro di indicatori di sopravvivenza di lungo periodo e di guarigione dopo un tumore, con l'obiettivo di fornire informazioni utili agli epidemiologi e ai pianificatori sanitari, ma anche agli oncologi e ai pazienti.

In particolare, vengono riportati risultati relativi ai seguenti indicatori:

- a. la proporzione di tutti i pazienti diagnosticati che si stima raggiungeranno lo stesso tasso di mortalità della popolazione generale (frazione di guarigione o *cure fraction*);
- b. il numero di anni dopo la diagnosi di tumore necessario a eliminare l'eccesso di mortalità legato al tumore (tempo per la guarigione o *time to cure*);
- c. la proporzione di pazienti prevalenti che, dopo la diagnosi, vivono da un numero di anni superiore al tempo per la guarigione e hanno già raggiunto tassi di mortalità simili alla popolazione generale (già guariti o *already cured*).

L'obiettivo di questo capitolo è di descrivere in dettaglio il *dataset* AIRTUM, il protocollo dello studio e i metodi statistici utilizzati nella presente Monografia.

IL DATABASE AIRTUM

La rete dei registri tumori (RT) italiani¹ include attualmente 40 registri tumori della popolazione generale e 5 registri specializzati. AIRTUM centralizza le informazioni raccolte dai singoli registri e ne supporta l'attività. I registri tumori generali includono una casistica relativa a quasi 30 milioni di persone, pari al 51% della popolazione Italiana (figura 1). AIRTUM verifica la qualità e la completezza dei dati e ne promuove l'uso per studi collaborativi di epidemiologia dei

AIMS

The aim of the present AIRTUM Report is to present up-to-date estimates and projections of the number of people living in Italy after a cancer diagnosis (i.e., complete prevalence). Complete cancer prevalence is defined as the proportion or the number of people in a population who have been diagnosed with cancer during their lifetime and who are still alive at a defined reference date, regardless of the date of diagnosis. It is a measure that depends on cancer incidence, survival, and on general mortality trends in the considered population. The ageing of the population increases the incidence of cancer, and, hence, it affects prevalence too. Cancer prevalence is related to the overall healthcare demand of cancer patients, including those who may be considered cured and who require only limited additional healthcare resources.

The present Report provides a wide spectrum of indicators of long-term survival and cure among cancer patients, in order to provide helpful information to epidemiologists and healthcare planners, as well as to oncologists and patients.

More specifically, we report results on:

- a. *the proportion of cancer cases, starting from the date of diagnosis, expected to reach the same death rates of the general population (cure fraction);*
 - b. *the number of years after cancer diagnosis necessary to eliminate excess mortality due to cancer (time to cure);*
 - c. *the proportion of prevalent cancer patients who have survived longer than "time to cure" and who have already reached the same death rates of the general population (already cured).*
- The aim of the present chapter is to describe the AIRTUM dataset, the study protocol, and the statistical methods used in the analyses.*

THE AIRTUM DATABASE

The Italian Network of Cancer Registries (CR)¹ at present includes 40 general population-based cancer registries and 5 specialised CR. AIRTUM collects data from cancer registries and supports their activity. General cancer registries comprehend almost 30 million people, 51% of the Italian population (figure 1). AIRTUM verifies data quality and completeness and uses data for collaborative studies on cancer epidemiology in Italy.²⁻⁵ Since 2005, AIRTUM has a central database, which stores data from all accredited cancer registries (i.e., CRs able to provide comparable high quality statistics on the incidence of cancer).⁶ The present report took advantage of the AIRTUM database updated to January 2014.

tumori.²⁻⁵ Dal 2005, il *dataset* centralizzato AIRTUM conserva i dati di tutti i registri tumori accreditati, cioè che hanno dimostrato di seguire procedure che soddisfino gli standard qualitativi richiesti a livello internazionale.⁶

La presente Monografia ha utilizzato il *database* AIRTUM aggiornato al gennaio 2014.

CRITERI DI INCLUSIONE

Al fine di stimare la prevalenza completa al 1.1.2010, sono stati applicati i seguenti criteri di inclusione dei registri:

1. disponibilità di almeno 5 anni consecutivi di incidenza;
2. ultimo anno di incidenza pari al 2007 o successivo;
3. follow-up completo dello stato in vita per almeno un anno successivamente all'ultimo anno di incidenza disponibile.

Per ciascun registro sono stati effettuati controlli relativi alla consistenza e alla completezza dei dati di incidenza e follow-up. Per i registri che hanno modificato le aree di copertura durante il periodo di osservazione (Provincia di Genova, Veneto, Romagna) sono state incluse le aree più estese in grado di fornire almeno 15 anni di osservazione (figura 2). Lo studio ha incluso 29 registri tumori italiani di popolazione, con una copertura di circa 21 milioni di persone, pari al 35% della popolazione italiana (figura 2 e tabella 1). Gli indicatori di completezza e qualità, usualmente utilizzati nell'ambito dei registri tumori italiani, sono stati ritenuti soddisfacenti.^{7,8}

Tra il 1976 e il 2009, nelle aree coperte dai registri tumori italiani inclusi nello studio sono stati diagnosticati 1.624.533 tumori maligni (tabella 2).

Nell'appendice 1 (pp. 134-135) viene mostrato il numero di casi incidenti nell'intero periodo considerato a seconda del tipo di tumore, del sesso e dell'età alla diagnosi dei pazienti. Tutti i casi di tumore presenti nel *database* AIRTUM sono stati codificati secondo la terza edizione della classificazione internazionale delle malattie (ICD-O-3). Sono stati inclusi tutti i tumori invasivi (ICD-10: C00-43, C45-96) diagnosticati in tutte le età, tranne quelli non melanomatosi della cute, perché per questi tumori la registrazione è considerata insufficientemente completa e accurata. Sono stati inclusi i tumori della vescica con un comportamento benigno, incerto o *in situ* (ICD-10: D09.0, D30.3, D41.4). I codici morfologici ICD-O-3 sono stati utilizzati al fine di definire specifici sottotipi di tumore, in particolare per definire neoplasie linfopoietiche clinicamente rilevanti e omogenee.^{9,10}

La tabella 2 mostra i 50 tipi o raggruppamenti di tumore inclusi nella presente Monografia. Le sedi rare o mal definite (ICD-10: C26, C39, C48, C55, C57, C58, C63, C74, C75, C76, C77, C80), circa il 3% di tutti i tumori diagnosticati nel periodo in esame, sono state incluse nel raggruppamento «Tutti i tumori».

Nel calcolare la prevalenza specifica per tipo di tumore, è stato considerato per ciascun individuo solo il primo tumore sviluppato in quella specifica sede. Tuttavia, i tumori primari multipli diagnosticati nella stessa persona in diffe-

INCLUSION CRITERIA

To provide an estimation of complete prevalence up to 1.1.2010, the following inclusion criteria for Cancer Registries were applied:

1. availability of at least five consecutive years of incidence;
2. last year of incidence equal to 2007 or later;
3. complete follow-up for at least one year after the last available year of incidence.

Controls on data consistency and completeness of incidence trends and follow-up were performed for each single registry dataset. In addition, for cancer registries that enlarged their catchment area during the period of interest, we included cases resident in the largest available area, provided that at least 15 years of observations were available (Genoa Province, Veneto, and Romagna) (figure 2).

Overall, 29 population-based Italian cancer registries were included in the study, covering approximately 21 million people (35% of the Italian population) (figure 2 and table 1). Routine indicators of data completeness and quality of Italian cancer registries were deemed satisfactory.^{7,8} Between 1976 and 2009, 1,624,533 malignant cancers cases were diagnosed in the areas covered by included CRs (table 2).

Incident cases in the overall period considered by sex, cancer type, area, and age are described in detail in appendix 1 (pp. 134-135).

All cases in the AIRTUM database were coded according to the third edition of the International Classification of Diseases for Oncology (ICD-O-3). Cases of all invasive cancers of all ages were included (ICD-10: C00-43, C45-96) with the exclusion

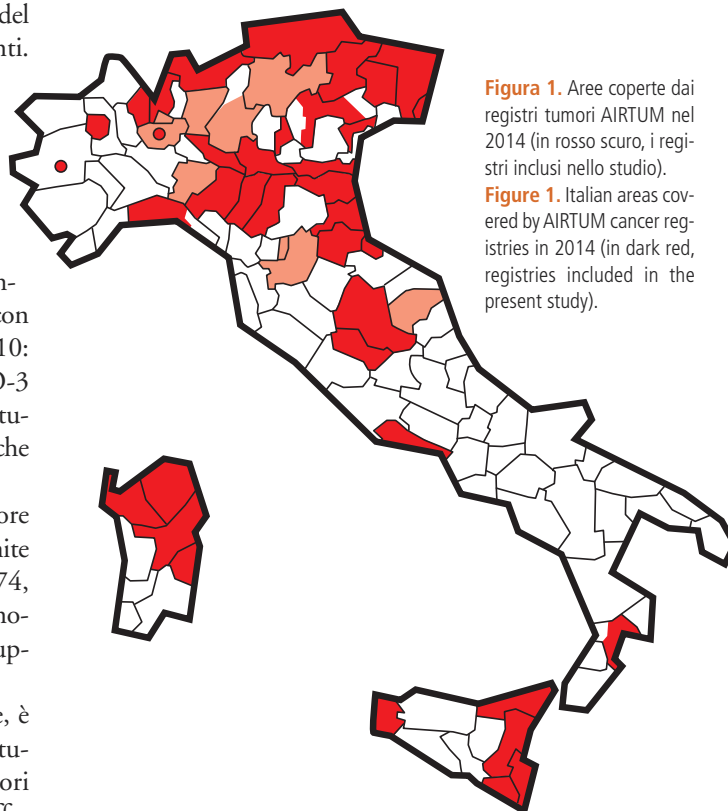


Figura 1. Aree coperte dai registri tumori AIRTUM nel 2014 (in rosso scuro, i registri inclusi nello studio).

Figure 1. Italian areas covered by AIRTUM cancer registries in 2014 (in dark red, registries included in the present study).

	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	2006	2007	2008	2009	2010	2011	2012	FOLLOW-UP		
NORTH-WEST	Biella																																					01/12/09		
	Como																																						31/12/12	
	Cremona																																						31/12/11	
	Genova																																						31/12/11	
	Mantova																																						30/09/13	
	Milano																																						31/01/10	
	Sondrio																																						31/03/12	
	Torino																																							31/01/09
	Varese																																							31/01/13
	NORTH-EAST	Alto Adige																																						31/03/13
Ferrara																																								31/12/11
Friuli Venezia Giulia																																							31/12/12	
Modena																																								31/12/11
Parma																																								31/12/12
Reggio Emilia																																								31/12/10
Romagna																																								31/12/12
Veneto																																							30/10/12	
CENTRE	Latina																																							31/12/11
	Umbria																																							31/12/11
SOUTH AND ISLANDS	Catania-Messina																																							31/12/11
	Catanzaro																																							31/12/08
	Napoli																																							31/12/10
	Nuoro																																							31/12/10
	Palermo																																							31/01/12
	Ragusa																																							31/12/11
	Salerno																																							31/12/12
	Sassari																																							31/12/10
	Siracusa																																							31/12/09
Trapani																																							31/03/10	

In bold: long term cancer registries included to estimate model-based incidence and survival.

Figura 2. Dati di incidenza disponibili per il periodo 1976-2012 e follow-up (in giallo) dei 29 registri tumori che hanno contribuito alla presente Monografia.

Figure 2. Available incidence data for the period 1976-2012 and follow-up (yellow) of the 29 cancer registries contributing to the present Report.

renti organi sono stati inclusi nelle analisi per ciascun tipo di neoplasia. Nel caso di stime riferite a raggruppamenti di tumori è stato incluso solo il primo tumore all'interno del raggruppamento stesso. Per questo motivo la somma dei pazienti prevalenti per i singoli tipi di tumore può essere superiore a quella mostrata nei raggruppamenti o nella categoria «Tutti i tumori».

PREVALENZA COMPLETA

La stima della prevalenza completa è stata ottenuta partendo dalla prevalenza osservata in ciascun registro, aggiustata attraverso gli indici di completezza che includono la frazione di casi prevalenti non osservati poiché diagnosticati prima dell'avvio della registrazione.¹¹

La prevalenza osservata viene calcolata direttamente dai dati di incidenza e stato in vita raccolti dai registri tumori ed è necessariamente incompleta, in quanto si riferisce al numero e alla proporzione di casi diagnosticati dopo l'inizio dell'attività di registrazione.^{2,11-12} Di conseguenza, la prevalenza osservata dai registri tumori italiani rappresenta una quota variabile della prevalenza completa, che aumenta con la durata del periodo di osservazione. Solo i registri con almeno 40-50 anni di osservazione possono virtualmente misurare con buona approssimazione la prevalenza completa. I primi registri tumori

of non-melanoma skin cancers, because registration for such cancers is likely to be less complete and accurate than for other cancer sites.

Urinary bladder cancers with benign or uncertain behaviour and in situ were included (ICD-10: D09.0, D30.3, D41.4). To define specific subtypes, ICD-O-3 morphology codes were used, in particular to define clinically relevant and homogeneous haemolymphopoietic neoplasms.^{9,10} The 50 cancer types or combinations included in the present Report are shown in table 2. Other rare or ill-defined cancers (ICD-10: C26, C39, C48, C55, C57, C58, C63, C74, C75, C76, C77, C80), approximately 3% of incident cancer cases in the whole considered period, were included in «all types» combination.

To estimate cancer-specific prevalence, we considered for each individual only one primary cancer occurring in that specific anatomic site. Nevertheless, multiple primary cancers in different organs diagnosed in the same person were computed in each site-specific analysis. For estimates of combinations of cancer types, only the first primary tumour in the group was included. Therefore, the sum of type-specific, prevalent patients was higher than the total number of prevalent subjects in the combination categories or in the «all types» category.

MACROAREA/ REGION	POPULATION	POPULATION IN AREAS COVERED BY CR	
	(No.)	(No.)	(%)
Piemonte	4 362 041	1 061 614	24
Valle d'Aosta	126 686	0	0
Lombardia	9 600 951	3 609 063	38
Liguria	1 576 443	860 444	55
NORTH-WEST	15 666 121	5 531 121	35
Trentino-Alto Adige	1 017 111	498 315	49
Veneto	4 841 933	2 111 955	44
Friuli Venezia Giulia	1 221 569	1 221 569	100
Emilia-Romagna	4 306 684	3 056 068	71
NORTH-EAST	11 387 297	6 887 907	60
Toscana	3 657 340	0	0
Umbria	880 202	880 202	100
Marche	1 540 272	0	0
Lazio	5 442 963	537 670	10
CENTRE	11 520 777	1 417 872	12
Abruzzo	1 307 778	0	0
Molise	315 536	0	0
Campania	5 758 375	1 663 199	29
Puglia	4 048 007	0	0
Basilicata	581 140	0	0
Calabria	1 966 336	234 124	12
Sicilia	4 997 429	4 103 417	82
Sardegna	1 641 347	685 828	42
SOUTH AND ISLANDS	20 615 948	6 686 568	32
ITALY	59 190 143	20 523 468	35

Tabella 1. Distribuzione della popolazione residente in Italia al 1.1.2010 per l'intero Paese e nelle aree coperte dai registri (CR) tumori inclusi nella presente Monografia, per regione e macroarea. Italia, 2013. (Fonte: ISTAT).¹³

Table 1. Distribution of the Italian resident population at 1.1.2010 in the whole Country and in areas covered by cancer registries (CR) included in the present Report, by region and macroarea. Italy, 2013. (Source: Italian National Institute of Statistics).¹³

italiani hanno iniziato l'attività dal 1976 e la durata di osservazione al 1.1.2010 variava da 5 a 34 anni.

Gli indici di completezza consentono di stimare la frazione di prevalenza non osservata nei dati dei registri. Questi indici variano, oltre che a seconda del periodo di attivazione del registro, in funzione del tipo di tumore, essendo dipendenti sia dall'incidenza sia dalla sopravvivenza passata. La prevalenza completa dei casi a età x è composta da tutti i casi incidenti a ciascuna età t ($t < x$) che sono sopravvissuti per $x-t$ anni, cioè fino all'età x . Un registro attivo da L anni potrà osservare solo i casi prevalenti con una durata della malattia inferiore a L .^{2,11} La prevalenza completa $Prev(x)$ può, quindi, essere scomposta in due componenti, la prima osservata (casi prevalenti che sopravvivono per durate da 0 a L) e la seconda non osservata (casi che sopravvivono per durate superiori a L), secondo la relazione seguente:

$$Prev(x) = Prev_L^{obs}(x) + Prev_L^{unobs}(x) \quad (1)$$

COMPLETE PREVALENCE

Complete prevalence estimation was based on registry-specific observed prevalence adjusted through completeness indices to include the fraction of unobserved prevalent cases who were diagnosed before the beginning of the registration.¹¹

Observed prevalence was directly computed from incidence and life status data collected by registries on their target population. The observed prevalence was necessarily incomplete, since it referred to the number and proportion of cases diagnosed after the start of the registry activity.^{2,11-12} Consequently, the observed prevalence of Italian registries represented a percentage of the complete prevalence increasing with length of observation time; only registries with 40-50 years of follow-up can detect a virtually complete observed prevalence. Italian registries started in different years from 1976 onwards and their observation periods ranged from 5 to 34 years at the index date of 1.1.2010.

The completeness indices allow to estimate the fraction of complete prevalence not observed in the recorded data. These indices vary not only by registration period length, but also by cancer type, since they depend on incidence and on survival as well. Complete prevalence at age x is composed of all incident cases diagnosed at any age t ($t < x$) surviving up to age x , thus for $x-t$ years. A population registry active since L years can observe only prevalent cases with a disease duration lower than L .^{2,11}

The expected complete prevalence $Prev(x)$ can, therefore, be broken down into two components, an observed (durations from 0 to L) and an unobserved (durations longer than L) one, according to the following relationship:

$$Prev(x) = Prev_L^{obs}(x) + Prev_L^{unobs}(x) \quad (1)$$

The completeness index is defined as the ratio between the expected limited duration prevalence (x) and the expected complete prevalence $Prev(x)$ for a specific cancer and observation time L , that is:

$$R_L(x) = \frac{Prev_L^{obs}(x)}{Prev(x)} = \frac{\sum_{t=x-L}^x I(t)RS(t,x-t)}{\sum_{t=0}^x I(t)RS(t,x-t)} \quad (2)$$

where:

$I(t)$ is the expected incidence of the disease at age t ;

$RS(t,x-t)$ is the expected relative survival at age x for a person who has been diagnosed with cancer at age t .

Both relative survival and incidence functions were estimated using data from the historical pool of eight cancer registries with registration periods ≥ 18 years. A systematic estimation of the completeness indices was performed by cancer type through the COMPREV software,¹⁴ using incidence and survival parameters calculated as described below.

L'indice di completezza viene definito, per ciascun tipo di tumore e tempo di osservazione L, come il rapporto tra i valori attesi della prevalenza osservata a durata limitata (x) e quella completa $Prev(x)$, cioè:

$$R_L(x) = \frac{Prev_L^{obs}(x)}{Prev(x)} = \frac{\sum_{t=x-L}^x I(t)RS(t,x-t)}{\sum_{t=0}^x I(t)RS(t,x-t)} \quad (2)$$

dove:

$I(t)$ rappresenta l'incidenza attesa della malattia all'età t; $RS(t,x-t)$ la sopravvivenza relativa attesa all'età x per una persona cui è stato diagnosticato il tumore all'età t. Sia la sopravvivenza relativa sia l'incidenza sono state stimate utilizzando il pool storico degli otto RT italiani con un periodo di osservazione ≥ 18 anni. La stima sistematica degli indici di completezza è stata effettuata tramite il software COMPREV,¹⁴ utilizzando i parametri di incidenza e sopravvivenza descritti in seguito.

MODELLI PER LA FUNZIONE DI INCIDENZA

La funzione di incidenza descrive il rischio di ammalarsi di tumore al variare dell'età per ciascuna coorte di nascita presente nella popolazione alla data indice della prevalenza. Nel presente studio è stato utilizzato, per ciascun tipo di tumore e sesso, un polinomio di sesto grado parametrizzato sull'età alla diagnosi:

$$I(t,k) = \left\{ 1 + e^{-\left[a_k + \sum_{i=1}^6 b_i \left(\frac{t-t_0}{m} \right)^i \right]} \right\}^{-1} \quad (3)$$

dove $I(t,k)$ rappresenta l'incidenza all'età t per la coorte di nascita k, inclusa come covariata categorica nel modello al fine di aggiustare i trend di rischio nelle diverse coorti di nascita. I parametri della funzione di incidenza sono stati stimati tramite un modello logistico (procedura logistica SAS) adattando i tassi grezzi di incidenza dei pazienti diagnosticati nel pool storico dei registri tra il 1985 e il 2009. I dati di incidenza sono stati raggruppati per tipo di tumore, sesso, quinquennio di età e coorte di nascita (<1899, 1900-1904, ..., 2005-2009). La bontà di adattamento dei diversi modelli di incidenza è stata valutata con il criterio di Akaike (AIC)¹⁵ e con il confronto grafico tra i tassi osservati e quelli stimati dal modello.^{2,16}

MODELLI PER LA FUNZIONE DI SOPRAVVIVENZA RELATIVA

La sopravvivenza relativa (RS) è il rapporto tra la probabilità di sopravvivere osservata in una coorte di pazienti oncologici e quella attesa in una popolazione confrontabile libera dalla malattia. Questa misura è basata sull'assunzione di cause competitive di morte indipendenti. La sopravvivenza attesa è stata calcolata a partire dai dati ufficiali di mortalità per tutte le cause¹³ appaiando i pazienti per area di residenza, sesso, età e anno di diagnosi. Sono stati esclusi dalle analisi i casi notificati solo attraverso il certificato di morte e quelli identificati incidentalmente durante le autopsie. La

CANCER TYPE	INCIDENT CASES	ICD-10 OR ICD-O-3-M
<i>All types but non melanoma</i>	1 624 533	
<i>Head and neck</i>	63 164	C00-14, C30-32
Tongue	7 410	C01-02
Mouth	8 297	C03-06
Salivary glands	3 469	C07-08
Oropharynx	5 283	C09-10
Nasopharynx	2 440	C11
Hypopharynx	3 601	C12-13
Oesophagus	12 938	C15
Stomach	85 362	C16
Small intestine	4 607	C17
<i>Colon, rectum, anus</i>	213 336	C18-21
Colon	147 030	C18
Rectum	61 559	C19-20
Anus	4 747	C21
Liver	53 790	C22
Gallbladder	22 850	C23-24
Pancreas	48 158	C25
Nasal cavities	2 346	C30-31
Larynx	25 101	C32
Lung, bronchus, trachea	193 889	C33-34
Thymus, heart, mediastinum	3 123	C37-38
Bone	3 125	C40-41
Skin melanoma	35 552	C43
Mesothelioma	6 327	C45
Kaposi sarcoma	3 700	C46
Connective tissue	8 722	C47,49
Female breast	216 291	C50
Vagina and vulva	6 452	C51-52
Cervix uteri	14 139	C53
Corpus uteri	35 643	C54
Ovary	25 656	C56
Penis	1 852	C60
Prostate	151 522	C61
Testis	7 799	C62
Kidney	50 718	C64-66,68
Bladder	113 720	C67, D09.0, D30.3, D41.4
Choroidal melanoma	1 614	C69, M8720-8790
Brain and central nervous system (CNS)	25 727	C70-72
Thyroid	32 327	C73
Hodgkin lymphoma	9 947	M9650-9667
<i>Non-Hodgkin lymphoma (NHL)</i>	57 564	C82-85,96
Chronic Lymphocytic Leukaemia/ Small Lymphocytic Lymphoma (CLL/SLL)	17 715	M9670,9823
NHL, Diffuse Large B-cell Lymphoma (DLBC)	15 570	M9678-9684
NHL, follicular	8 206	M9675,9690-9698
Multiple myeloma (plasma cell)	22 600	M9731-9734
<i>Leukaemia</i>	42 206	C91-95
Precursor cell acute lymphoblastic lymphoma/ leukaemia	4 530	M9727-9729, 9835-9837
Acute myeloid leukaemia	12 411	M9840,9861, 9866-9867, 9870-9874, 9891-9931
Chronic myeloid leukaemia	5 200	M9863,9875

Tabella 2. Tumori presi in considerazione in questa Monografia e relativi casi incidenti. In corsivo, i raggruppamenti di tumori.

Table 2. Cancer types considered in this Report and respective incident cases. In italics, combinations of cancer types.

RS è stata calcolata con il metodo di coorte e l'approccio Ederer II¹⁷ utilizzando il programma SEER*Stat.

La RS è stata stimata utilizzando le informazioni dei registri tumori con almeno 18 anni di osservazione (pool storico) (figura 2). Questi registri includevano il 48% di tutti i casi incidenti che hanno contribuito allo studio. Le RS di lungo periodo, fino a un massimo di 25 anni dalla diagnosi, sono state calcolate dall'insieme dei casi diagnosticati nel periodo 1985-2008 e seguiti fino al 2009.

Le RS sono state stimate per tipo di tumore, sesso, età alla diagnosi (0-14, 15-44, 45-59, 60-74 e ≥ 75 anni) e periodo di diagnosi (in trienni dal 1985-1987 al 2006-2008).

La funzione di sopravvivenza relativa è stata parametrizzata attraverso un modello misto con ipotesi di guarigione. I modelli misti consentono di stimare la frazione di guarigione (C , *cure fraction*), cioè la proporzione di pazienti che hanno la stessa attesa di vita (o tasso di mortalità) della popolazione generale, e la distribuzione del tempo alla morte per i pazienti che non guariranno ($1-C$).¹⁸

Per costruire il modello di sopravvivenza relativa cumulata per i pazienti a età x diagnosticati a età t nell'anno y si è usata una distribuzione Weibull:

$$RS(x, t, y) = [C + (1 - C) \exp(-(\lambda(x - t))^\gamma)]^{\exp[\beta_1(t - t_0) + \beta_2(y - y_0)]} \quad (4)$$

dove:

$x-t$ è la durata del follow-up;

λ e γ sono, rispettivamente, i parametri di scala e forma della distribuzione Weibull di mortalità dei pazienti che non guariranno;

β_1 e β_2 modulano l'effetto di età e anno alla diagnosi sulla sopravvivenza relativa dei pazienti.

Si ipotizza che età e anno di diagnosi siano variabili continue. Sono stati utilizzati come riferimento l'età mediana dei pazienti ($t_0 = 65$ anni) e l'anno centrale del periodo di osservazione ($y_0 = 1995$).

Il modello (4) è stato utilizzato per tutti i tipi di tumore, a eccezione dei tumori di mammella, tiroide e delle neoplasie emolinfopoietiche, per i quali si è ottenuto un migliore adattamento dei modelli di sopravvivenza utilizzando modelli misti di cura stratificati per età, con una distribuzione di sopravvivenza di tipo esponenziale.

La bontà di adattamento dei modelli di RS è stata valutata attraverso test di massima verosimiglianza e con il confronto grafico tra curve di sopravvivenza osservate e stimate dai modelli.¹⁶ I valori osservati e stimati sono risultati coerenti per ciascun tipo di tumore, sesso e gruppo di età.^{2,19} Per esempio, per la coorte di pazienti diagnosticati nel periodo 1994-1996 l'adattamento dei valori di RS modellizzati con quelli empiricamente osservati fino a 15 anni dopo la diagnosi è stato ottimo per ciascun tipo di tumore, sesso ed età, in particolare per i tumori più frequenti (mammella, polmone, colon retto, vescica e prostata).

INCIDENCE MODEL

The incidence function describes the relationship between age and risk of being diagnosed with cancer, as measured along the life span of each birth cohort present in the population at the prevalence date. In the present study, a sixth degree polynomial on age was used for each cancer type and sex:

$$I(t, k) = \left\{ 1 + e^{-\left[(a_k + \sum_{i=1}^6 b_i \left(\frac{t-t_0}{m} \right)^i \right]} \right\}^{-1} \quad (3)$$

where $I(t, k)$ is incidence at age t for birth cohort k .

The birth cohort covariate was included as a categorical variable to adjust for risk trends across different birth cohorts.

Incidence function parameters were estimated through the SAS logistic procedure by fitting crude incidence rates of patients registered between 1985 and 2009 in the historical pool. Incidence data were categorised according to cancer type, sex, five-year age, and birth cohort (<1899, 1900-1904, ..., 2005-2009). The goodness of fit of the various incidence models was assessed by Akaike Information Criterion (AIC),¹⁵ as well as by visual comparison between estimated and observed rates.^{2,16}

RELATIVE SURVIVAL MODEL

Relative Survival (RS) is the ratio of the observed survival probability in the cohort of cancer patients to the expected survival probability in a comparable set of cancer free individuals. This formulation is based on the assumption of independent competing causes of death. Expected survival was computed from all causes official mortality data¹³ by matching individuals by Cancer Registry area, sex, age, and year at diagnosis. Death certificates only and cases incidentally diagnosed at autopsy were excluded from the analysis. RS was calculated using the cohort method and the Ederer II approach¹⁷ by means of the SEER*Stat software.

RS was estimated for the restricted pool of eight cancer registries (historical pool), providing incidence data for a period of at least 18 years (figure 2). These registries included 48% of all incident cases contributing to the study. Long-term RS up to maximum 25 years after diagnosis was calculated by pooling cases diagnosed in 1985-2008 and followed-up until 2009. RS was estimated by cancer type, sex, age at diagnosis (0-14, 15-44, 45-59, 60-74, and ≥ 75 years), and period of diagnosis (in three-year periods from 1985-1987 to 2006-2008).

Model-based relative survival was calculated using mixture cure models, which estimate the cure fraction C , i.e., the proportion of cured patients who have the same life expectancy (or death rates) of the general population, and time to death distribution of "non-cured" patients ($1-C$).¹⁸

The following Weibull mixture cure model was used to model, by cancer type and sex, the cumulative RS probability of patients at age x diagnosed at age t and year y :

$$RS(x, t, y) = [C + (1 - C) \exp(-(\lambda(x - t))^\gamma)]^{\exp[\beta_1(t - t_0) + \beta_2(y - y_0)]} \quad (4)$$

VALIDAZIONE DELL'INDICE DI COMPLETEZZA

Per validare gli indici di completezza è stata utilizzata la prevalenza osservata nei registri di Varese (34 anni di osservazione) e Parma (32 anni). Per entrambi è stata calcolata la prevalenza a 5, 10, 15 e 20 anni, alle quali sono stati applicati gli indici di completezza per stimare la prevalenza fino alla massima durata di osservazione. Per esempio, per il registro di Varese, alle prevalenze a 5, 10, 15 e 20 anni sono stati applicati gli indici di completezza per calcolare la prevalenza a 34 anni. La validazione è stata condotta per i tumori con almeno 50 casi prevalenti per 100.000 abitanti nel periodo più lungo di osservazione.

Come prevedibile, le differenze tra valori osservati e stimati di prevalenza sono più alte per le stime ottenute usando solo i cinque anni più recenti di registrazione e diminuiscono all'aumentare del periodo di osservazione. Per il registro di Parma, le differenze medie variano, in termini assoluti, dall'11% per le stime basate su 5 anni, al 3% utilizzando quelle a 20 anni. Le stesse differenze variano dall'8% al 3% per il registro di Varese. Per l'insieme di tutti i tumori (esclusi i non melanomatosi della cute) in entrambi i registri, le differenze tra prevalenza osservata e stimata a partire da 10, 15 e 20 anni di osservazione erano inferiori al 4%.

INDICATORI DELLA SOPRAVVIVENZA DI LUNGO PERIODO E DI GUARIGIONE

Gli indicatori di guarigione dopo un tumore sono stati ottenuti utilizzando la sopravvivenza relativa calcolata dai modelli descritti e la sopravvivenza relativa condizionata (CRS), cioè la probabilità cumulativa che il paziente ha di sopravvivere ulteriori y anni, dato che è già sopravvissuto x anni. Questa è stata calcolata attraverso la relazione seguente:

$$CRS_{x,y} = RS_{x+y} / RS_x \quad (5)$$

La guarigione in termini statistici è osservabile per la maggior parte dei tumori. Non tutti i pazienti neoplastici moriranno in conseguenza della loro malattia e si possono definire guariti coloro che raggiungono gli stessi tassi di mortalità della popolazione generale (figura 3). Ciò significa che durante il periodo di osservazione questi pazienti rimarranno vivi o moriranno per altre cause.^{19,20}

La quota di pazienti guariti, cioè la *cure fraction*, è data dal valore raggiunto dalla curva di RS quando nel lungo periodo si appiattisce su un valore costante (figura 3). Al contempo, la CRS tenderà al 100% (figura 4), ovvero nessun eccesso di mortalità sarà dovuto al tumore.¹⁸⁻²⁴

La *cure fraction* e il tempo per la guarigione sono calcolati per tipo tumorale e sesso. Le stime per classe d'età sono state calcolate usando l'età media in ciascuna classe d'età:

- 35 anni per la classe d'età 0-44 anni per tutte le sedi, eccetto osso, tessuto connettivo, linfoma di Hodgkin, leucemie, leucemia mieloide acuta ed encefalo, dove l'età media è 25 anni; per le leucemie linfatiche acute, dove gli indicatori sono calcolati a 12 anni;
- 53 anni per la classe 45-59 anni;

where:

$x-t$ is the follow-up time;

λ and γ are, respectively, the scale and shape parameters of the Weibull distribution used to assess mortality of fatal cases;

β_1 and β_2 express the effects of "age at diagnosis" and "year of diagnosis", respectively, on the relative survival function of all patients.

In this model, age and period at diagnosis are assumed to be continuous variables. The median age of cases ($t_0 = 65$ years) and the central year of the observation period ($y_0 = 1995$) were used as reference. Model (4) was used for all cancer types except for breast and thyroid cancers, and for all haemolymphopoietic neoplasms.

For these tumours, a linear dependence was not suited to model survival age patterns and a better fit was obtained stratifying by age class and using mixture cure models with exponential survival distribution. The goodness of fit of model-based RS was assessed by likelihood ratio test and by visual comparison,¹⁶ for each cancer type, sex, and age group,^{1,19} showing consistent results. For the cohort diagnosed in 1994-1996, model-based RS fitted closely the empirical RS values up to 15 years after diagnosis, for each cancer type, sex, and age group, in particular for common cancer types (i.e., breast, lung, colon rectum, bladder, prostate).

VALIDATION OF COMPLETENESS INDICES

In order to validate the completeness indices, observed prevalence from the registries of Varese (34 years of observation) and Parma (32 years) was used. For each of these registries, we calculated truncated limited-duration prevalence at 5, 10, 15, and 20 years; then applied the corresponding completeness indices to estimate prevalence for the maximum observation period. For the Varese registry, for instance, we estimated 34-year prevalence by completing 5, 10, 15, and 20 year observed prevalence and compared it with observed 34-year prevalence. This validation was performed for all cancer types with at least 50 prevalent cases per 100,000 inhabitants in the longest period of observation.

As expected, the relative percent differences decreased with increasing observation length, being highest for estimates derived from only five years registration. The average absolute differences on all analysed cancer types ranged between 11% (5 years) and 3% (20 years) for Parma registry, and between 8% and 3% for Varese registry. For all malignant cancers (except non-melanoma skin cancers), the absolute relative differences were below 4% in both registries, when prevalence was truncated at 10, 15, or 20 years.

INDICATORS OF LONG-TERM SURVIVAL AND CURE

Indicators of the cure of cancer were derived from model-based relative survival and from conditional relative survival (CRS), i.e., the cumulative probability of surviving additional y years, given that patients already survived x years. This is derived from the following ratio:

$$CRS_{x,y} = RS_{x+y} / RS_x \quad (5)$$

Cure is observable, in statistical terms, for most cancer types. Not

Figura 3. Frazione di guarigione: proporzione di tutti i pazienti diagnosticati che raggiungeranno lo stesso tasso di mortalità della popolazione generale. Tale proporzione (linea tratteggiata) è pari al plateau raggiunto dalla curva della sopravvivenza relativa (43% nell'esempio basato sui dati del cancro coloretale negli uomini).

Figure 3. Cure fraction: proportion of cancer patients who are expected to reach the same death rates of the general population and will not die as a result of their cancer.

This proportion (dashed line) equals the plateau reached by the relative survival curve (43% in this example based on colorectal cancer data for Italian men).

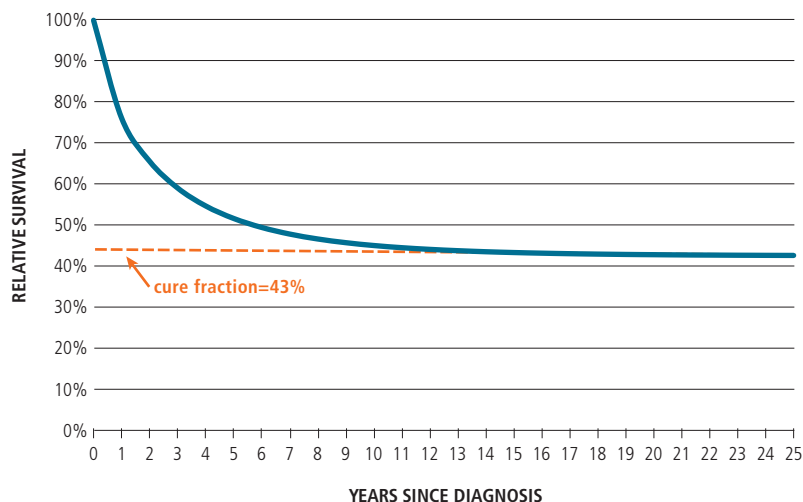


Figura 4. Tempo per la guarigione: numero di anni necessario, dopo la diagnosi di tumore, a eliminare l'eccesso di mortalità legato al tumore stesso. È misurato come il tempo necessario per raggiungere una sopravvivenza relativa condizionata (CRS) a 5 anni >95% (9 anni nell'esempio basato sui dati del cancro coloretale negli uomini). Il tasso di mortalità dei pazienti oncologici è lo stesso della popolazione generale quando la CRS a 5 anni raggiunge il 100%. Quando la CRS a 5 anni (linea blu) diventa >95%, l'eccesso di mortalità può essere classificato come trascurabile.

Figure 4. Time to cure: time span after cancer diagnosis necessary to eliminate excess mortality due to cancer. It is measured as the time necessary to reach a 5-year Conditional Relative Survival (CRS) >95% (9 years in this example based on colorectal cancer data for Italian men). Death rates of cancer patients are the same of those in the general population when five-year CRS reaches 100%. When five-year CRS (blue line) becomes >95%, excess mortality can be classified as "hardly any".

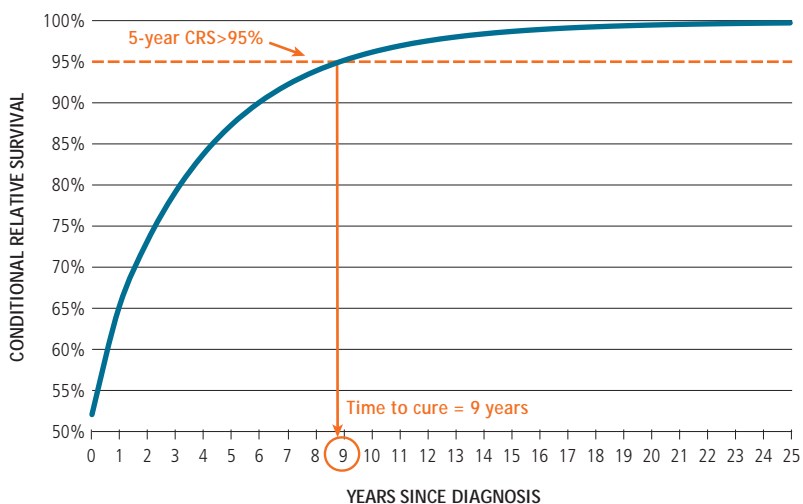


Figura 5. Distribuzione dei casi prevalenti per tempo dalla diagnosi in anni.

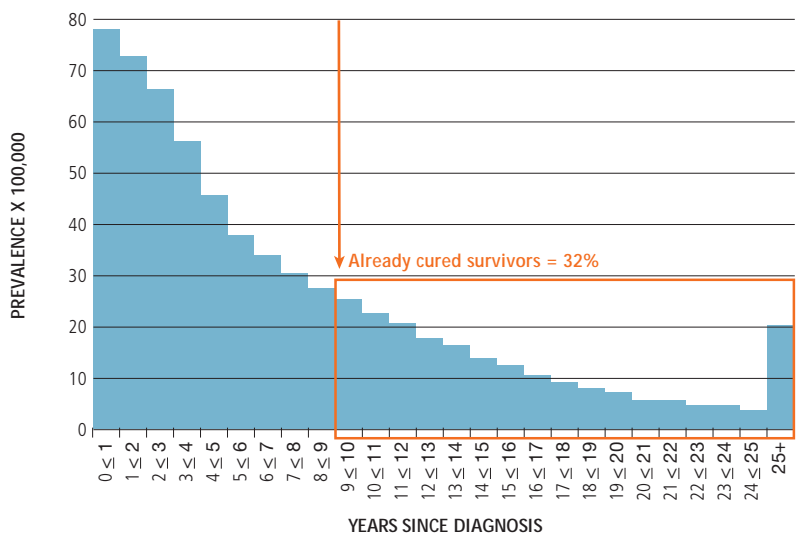
Già guariti: proporzione di pazienti prevalenti che vivono da un numero di anni superiore al tempo per la guarigione.

Il rettangolo include i pazienti già guariti, cioè coloro che sono sopravvissuti per un numero di anni superiore al tempo di guarigione (32% nell'esempio basato sui dati del cancro coloretale negli uomini).

Figure 5. Distribution of prevalent cases by time since diagnosis in years.

Already cured survivors: Patients who have survived longer than time to cure.

The rectangle encloses the already cured patients, i.e., those who have survived since more than time to cure (32% years in this example based on colorectal cancer data for Italian men).



- 72 anni per la classe 60-74 anni;
- 82 anni per la classe ≥ 75 anni.

Tutti gli indicatori di guarigione sono calcolati per l'insieme degli 11 RT con incidenza al 2009 e almeno 7 anni di registrazione (37% dell'intera casistica) in cui è possibile calcolare la prevalenza per singolo anno dalla diagnosi (figura 2).

TEMPO PER LA GUARIGIONE

Il tempo per la guarigione può essere stimato assumendo che la guarigione si raggiunga quando la CRS si avvicina al 100% (per esempio, diventa maggiore del 90% o del 95%).^{19,25-27} Nella presente Monografia il tempo per la guarigione è stato definito come il numero di anni necessari affinché la CRS a 5 anni, calcolata attraverso i modelli, diventi maggiore del 95%. Si assume, cioè, che un eccesso di mortalità inferiore al 5% nei successivi 5 anni sia clinicamente ed epidemiologicamente trascurabile.^{19,26}

Nell'esempio (figura 4) il tempo per la guarigione viene raggiunto dopo 9 anni dalla diagnosi. Sono stati proposti anche altri livelli (CRS a 5 anni $>90\%$)^{19,26} e indicatori (CRS a 10 anni)¹⁹ del tempo per la guarigione. Tuttavia, la combinazione scelta (CRS a 5 anni $>95\%$), oltre a essere clinicamente rilevante, consente il confronto con stime effettuate in altri Paesi.²⁵⁻³³

PROPORZIONE DI PAZIENTI GIÀ GUARITI

Abbiamo definito come guariti dal tumore i pazienti prevalenti ammalatisi da un numero di anni superiore al tempo per la guarigione.¹⁹ In questo Rapporto vengono definiti "già guariti". Oltre a questi pazienti, una quota ulteriore di persone guariranno: sono coloro la cui diagnosi è avvenuta da un numero di anni inferiore al tempo per la guarigione e non hanno ancora raggiunto questa soglia.

La figura 5 mostra un esempio in cui il tempo per la guarigione viene raggiunto in circa 9 anni e l'area nel rettangolo mostra la proporzione di pazienti già guariti.

Le stime del tempo per la guarigione si basano sulla CRS calcolata per classe di età alla diagnosi. La prevalenza viene calcolata, invece, per classi di età alla data indice di prevalenza (1.1.2010). Questa tecnica, che implica un'ipotesi semplificatrice, può causare una sottostima dell'indicatore, in quanto il tempo necessario a raggiungere un eccesso di mortalità trascurabile aumenta per la maggior parte dei tipi di tumore ed è stato applicato alla prevalenza completa calcolata per classi di età più avanzate.

Ciascuna delle schede presentate (vd. capitolo 4 «Schede specifiche per tumore», pp. 43-93) mostra risultati che si riferiscono all'anno centrale di diagnosi nella popolazione (1995) ed è importante sottolineare che tutte le stime di sopravvivenza di lungo periodo e di guarigione rappresentano la sopravvivenza media di casistiche adeguatamente numerose di pazienti (stime probabilistiche a livello di popolazione) piuttosto che caratteristiche prognostiche individuali.

Certi raggruppamenti di tumori includono sottotipi con sopravvivenze molto eterogenee sia a breve, sia a lungo termine.

all patients die of cancer and we can define as cured those patients who reach the same mortality rates of the general population (figure 3). During the period of observation, these patients will remain alive or will not die as a result of their neoplasm.^{19,20}

When this occurs, the RS curve reaches a plateau that represents both the long term RS value and the proportion of cured patients, i.e., the cure fraction (figure 3). At the same time, CRS approaches 100% (figure 4), meaning no excess mortality due to cancer.¹⁸⁻²⁴

Cure fraction and time to cure are calculated by cancer type and sex. Estimates were also calculated by age group using the mean age in the group as reference for the estimate. This was:

- 35 years for 0-44 age class, except for bone, connective tissue, Hodgkin lymphoma, leukaemia, acute myeloid leukaemia, and brain, where the mean age was 25 years, and acute lymphoid leukaemia, where the mean age was 12 years;
- 53 years for 45-59 age class;
- 72 years for 60-74 age class;
- 82 years for ≥ 75 age class.

All cure indicators were calculated for the 11 registries with last available year of incidence equal to 2009 and at least 7 years of incidence (37% of total cases included in the analysis) for which it is possible to calculate the prevalence by single year since diagnosis (figure 2).

TIME TO CURE

Time to cure can be measured by means of model-based CRS curves assuming that cure is reached when five-year CRS approaches 100%, for instance, it becomes higher than 90% or 95%.^{19,25-27} In this Report, time to cure is defined as the number of years after which the model-based 5-year CRS becomes higher than 95%, thus assuming the residual 5% excess mortality to be clinically or epidemiologically not relevant (hardly any).^{19,26} In the example (figure 4), time to cure is reached after 9 years. Other thresholds (e.g., 90%)^{19,26} or other indicators (10-year CRS)¹⁹ can be considered. Nevertheless, the combination used (5-year CRS $>95\%$) is not only clinically relevant, but it also allows comparability with other Countries.²⁵⁻³³

PROPORTION OF ALREADY CURED PATIENTS

Prevalent cases who survived at the index date for a period longer than time to cure were considered as cured from cancer.¹⁹ In this Report, they are defined as "already cured". Moreover, an additional proportion of patients will be cured. These are the patients whose diagnosis has occurred since less than time to cure and they have not already reached this threshold.

In figure 5 an example is reported where time to cure is 9 years and the proportion of already cured prevalent cases is given by the area in the rectangle.

Estimates of time to cure are based on CRS that is calculated by age class at diagnosis. It should be noted that prevalence is computed by age class at the prevalence index date (1.1.2010). This method, based on a simplified assumption, could lead to a slight underestimation of this indicator since the time to reach negligible excess mortality increased with age for most cancer

La stima del tempo per la guarigione per questi raggruppamenti tende a rappresentare il valore dei sottotipi specifici il cui rischio di morte si mantiene più a lungo superiore rispetto alla popolazione generale. Per questo motivo i parametri di guarigione vengono mostrati solo per i sottotipi specifici nel caso delle leucemie. Per i sottotipi di linfomi, invece, l'instabilità delle stime di sopravvivenza di lungo periodo non consente di effettuare stime attendibili dei parametri di guarigione.

PROIEZIONI AL 2010

Per i RT con dati di incidenza anteriori al 2009, la prevalenza completa è stata stimata per gli ultimi tre anni disponibili. Per ciascun RT, tipo di tumore, sesso ed età sono stati calcolati il numero di casi prevalenti al 1.1.2010 attraverso modelli di regressione lineari che consideravano l'anno di calendario come variabile indipendente. La validazione di questa procedura è stata effettuata utilizzando i RT di Parma e Varese e simulando un troncamento dell'incidenza al 2007. Il confronto tra i valori di prevalenza osservata usando i dati di incidenza al 2009 e quelli proiettati con troncamento al 2007 non hanno mostrato alcuna differenza rilevante.

PROIEZIONI AL 2015

La prevalenza completa al 2015 è stata calcolata in modo simile attraverso semplici modelli di regressione lineare, con l'ipotesi che la prevalenza completa al 2015 (CP_{2015}) possa essere prevista con sufficiente grado di approssimazione da una funzione lineare in cui il termine di regressione è rappresentato dall'anno di calendario, come nella relazione seguente:

$$CP_{2015}(y) = \beta_0 + \beta_1 \cdot y + \varepsilon \quad (6)$$

dove:

β_0 rappresenta l'intercetta;

β_1 è la variazione media annua stimata dalla prevalenza completa negli ultimi tre anni disponibili per ciascun RT. Questa assunzione semplificatrice, trend lineari ad andamenti costanti, può essere considerata ragionevole per proiezioni a breve termine (per esempio, a 5 anni), ma per proiezioni a più lungo termine resta discutibile e non è verificabile.

Sono state esplorate altre assunzioni sull'andamento della prevalenza (per esempio, modelli log-lineari). Questi modelli hanno mostrato risultati consistenti con quelli ottenuti con i modelli lineari per i tipi di tumore più frequenti, ma si sono mostrati più instabili per quelli più rari.

Metodi statistici più complessi sono disponibili per effettuare proiezioni di prevalenza,^{34,35} ma tali stime di lungo periodo vanno oltre gli obiettivi della presente Monografia. Al fine di ridurre l'instabilità delle proiezioni per i RT più recenti (con meno di 7 anni di registrazione), il tasso di incremento annuo (β_1) è stato calcolato utilizzando i RT di lungo periodo presenti nelle macroaree corrispondenti.

Tutte le proiezioni sono state effettuate per tipo di tumore, sesso e gruppo di età. La figura 6 mostra il diagramma di flusso che sintetizza dati e programmi utilizzati nelle diverse

types, and it was applied to complete prevalence of more advanced (reference) ages.

All cancer-specific data sheets report results for the central year of diagnosis (1995) and it should be noted that all estimates of long-term survival and cancer cure estimates reflected the average survival time of large groups of people (i.e., a population) rather than an individual prognosis.

Particular combinations of cancers (e.g., leukaemias) include subtypes showing a huge short- and long-term survival heterogeneity. The estimates of time to cure for these combinations tend to represent survival of subtypes showing a persisting higher death risk, in comparison with the general population. As a consequence, indicators of cure were calculated only for specific leukaemia subtypes. Conversely, they were calculated only for all lymphomas, due to the instability of long-term survival estimates for lymphoma subtypes.

PROJECTIONS AT 2010

For CRs with missing incidence data in 2009, complete prevalence was estimated at the last three calendar years available at index dates. Thus, for each registry, cancer type, sex, and age, the number of prevalent cases was projected to the common date of 1.1.2010 by means of a linear regression model with the calendar year as an independent variable.

This projection procedure was validated with data from the Parma and Varese CRs by simulating incidence data truncation at 2007, and comparing projected and observed prevalence at 2009. Negligible differences emerged.

PROJECTIONS FOR 2015

Complete prevalence for 2015 was calculated by a simple linear regression models. It has been supposed that complete prevalence at 2015 (CP_{2015}) can be predicted by a linear function of a regressor variable calendar year (y).

For each CR, cancer type, sex, and age group, have been calculated as:

$$CP_{2015}(y) = \beta_0 + \beta_1 \cdot y + \varepsilon \quad (6)$$

where:

β_0 is the intercept;

β_1 the slope calculated using the complete prevalence for the last three calendar years available before the index dates.

This simplified assumption (linear and constant trend) may not be valid for long-term projections, but it is reasonable for short-term (e.g., 5-year) ones.

Other assumptions (e.g., log-linear models) were explored, showing consistent results for common cancer types, but unstable projections for the rarest.

More complex statistical methods to extend projection are available,^{34,35} but long-term projections are beyond the aim of the present Report.

In order to reduce instability of projections for recently established registries (<7 years of observation), area-specific, rather than registry specific, slope parameters (β_1) were used. Registries

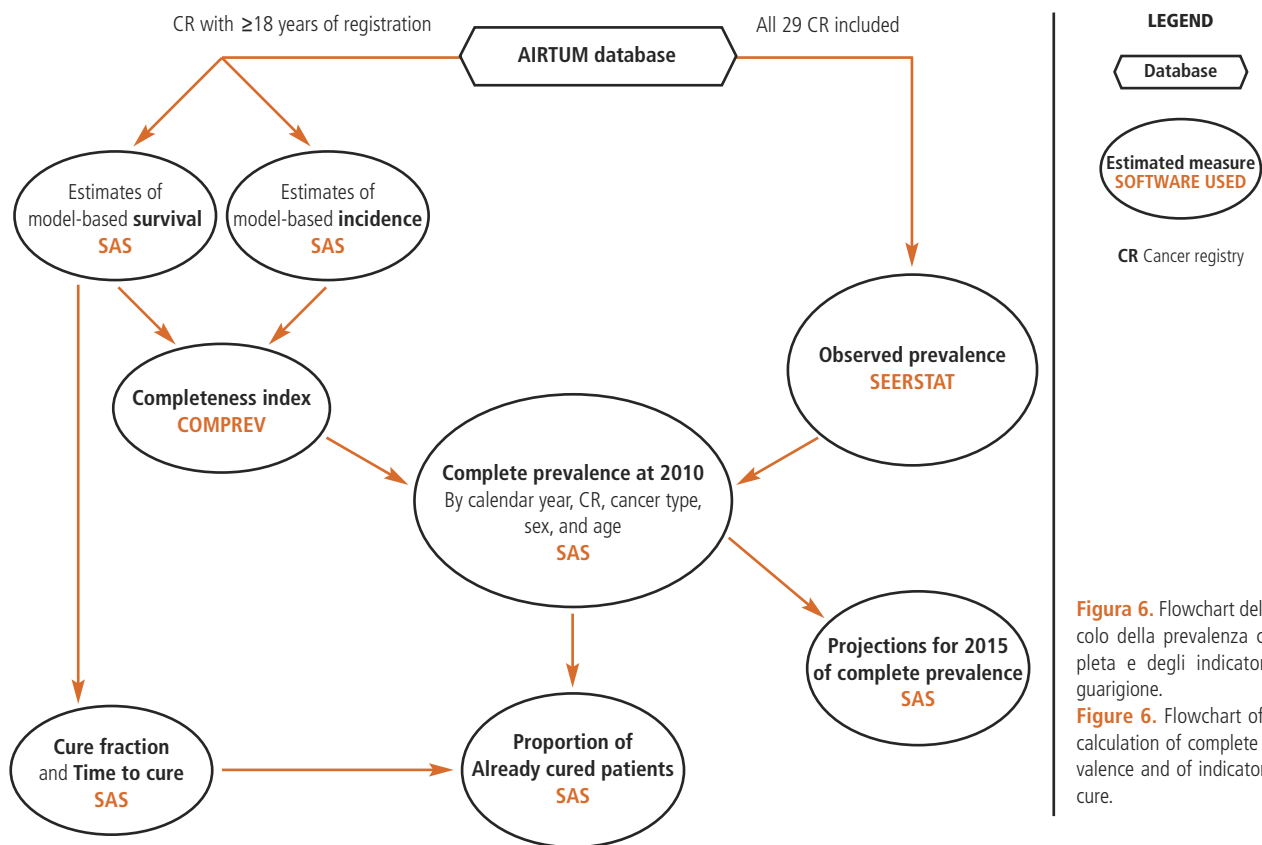


Figura 6. Flowchart del calcolo della prevalenza completa e degli indicatori di guarigione.
Figure 6. Flowchart of the calculation of complete prevalence and of indicators of cure.

fasi necessarie per il calcolo della prevalenza sin qui descritte. Certi raggruppamenti di tumori (es.: tutte le leucemie) includono sottotipi con sopravvivenze molto eterogenee sia a breve che a lungo termine. La stima del tempo alla guarigione per questi raggruppamenti tende a rappresentare il valore dei sottotipi specifici il cui rischio di morte si mantiene più a lungo superiore rispetto alla popolazione generale. Per questo motivo i parametri di guarigione vengono mostrati solo per i sottotipi specifici nel caso delle leucemie. Per i sottotipi di linfomi, invece, l'instabilità delle stime di sopravvivenza di lungo periodo non consente di effettuare stime attendibili dei parametri di guarigione.

STIME DI PREVALENZA NAZIONALI E REGIONALI

La stima del numero di persone vive in Italia dopo una diagnosi di tumore è stata ottenuta come somma del numero di casi prevalenti in ciascuna delle quattro macroaree nelle quali l'Italia è stata suddivisa (figura 1). Questi ultimi sono stati calcolati moltiplicando le proporzioni specifiche per età di casi prevalenti in ciascuna macroarea per la popolazione delle aree al 1.1.2010.¹⁴

Le stime nazionali di prevalenza completa ottenute come somma delle macroaree sono diverse da quelle stimate per l'insieme dei RT, perché i livelli di prevalenza e copertura di registrazione possono essere leggermente diversi tra le macroaree italiane.

Le stime regionali di prevalenza per tipo di tumore e sesso sono state calcolate in modo analogo anche per le regioni

with longer observation period were pooled to compute these area-specific slopes. All projections were achieved by cancer type, sex, and age group.

Figure 6 shows the flow chart of the calculation of complete prevalence and other indicators above described.

NATIONAL AND REGIONAL ESTIMATES OF PREVALENCE

The estimated complete number of prevalent patients by cancer type and sex in Italy was derived by summing up the number of prevalent cases in each of the four Italian macroareas (figure 1). These numbers were derived by multiplying age-specific complete prevalence proportions by macroarea to the corresponding resident population at 1.1.2010.¹⁴

The national complete prevalence by macroarea, derived by pooling registries present in each of the four areas, necessarily diverged from the prevalence proportion estimated for the Italian pool of CRs, because prevalence levels and registration coverage among areas were slightly different.

The regional estimated number of complete prevalent patients by cancer type and sex in regions with partial registration coverage was derived using the same approach, i.e., by multiplying the type-, sex-, age-, and macroarea-specific complete prevalence proportions by the corresponding regional population at 1.1.2010.¹⁴ For projections to 2015 (see previous paragraph), the same approach was applied by using regional populations provided by the Italian National Institute of Statistics (ISTAT).¹⁴ National and regional estimates assumed that populations

parzialmente coperte da RT, ovvero moltiplicando i tassi nella macroarea per la popolazione regionale corrispondente al 1.1.2010.¹⁴

Per le proiezioni al 2015, descritte al paragrafo precedente, è stato usato lo stesso approccio utilizzando le stime di popolazione regionale fornite dall'ISTAT.¹⁶

Sia le stime nazionali sia quelle regionali si basano sull'assunzione che la popolazione che vive nelle aree coperte dai RT sia rappresentativa di quella delle intere macroaree, comprese quelle non incluse nei RT, e del Paese. Le appendici 2 e 3 (pp 136-139) mostrano, rispettivamente, le popolazioni delle aree coperte e delle intere regioni e macroaree.

living in areas covered by a cancer registry are representative of population living in all macroareas, including those not covered by registries, or in the whole Country. Appendices 2 and 3 (pp. 136-139) show the populations used in estimates for CR areas and overall regions, respectively.

BIBLIOGRAFIA/REFERENCES

1. www.registri-tumori.it
2. AIRTUM Working Group. Italian cancer figures, report 2010: Cancer prevalence in Italy. Patients living with cancer, long-term survivors and cured patients. *Epidemiol Prev* 2010;34(5-6) Suppl 2:1-188.
3. AIRTUM Working Group. Italian cancer figures, report 2011: Survival of cancer patients in Italy. *Epidemiol Prev* 2011;35(5-6) Suppl3:1-200. Available at: <http://www.registri-tumori.it/cms/Rapp2011>
4. AIRTUM Working Group; CCM; AIEOP Working Group. Italian cancer figures, report 2012: Cancer in children and adolescents. *Epidemiol Prev* 2013;37(1) Suppl 1:1-225.
5. AIRTUM Working Group. Italian cancer figures, report 2013: Multiple tumours. *Epidemiol Prev* 2013;37(4-5) Suppl 1:1-152.
6. Protocollo operativo per la gestione della Banca Dati AIRTUM – 2.0 (Database AIRTUM Protocol). Available at: <http://www.registri-tumori.it/cms/files/2010.pdf>
7. Curado MP, Edwards B, Shin HR et al (eds). *Cancer Incidence in Five Continents. Vol. IX*. IARC Scientific Publication No. 160. Lyon, IARC, 2007. Available at: <http://www.iarc.fr/en/publications/pdfs-online/epi/sp160/Ci5vol9-A.pdf>
8. Forman D, Bray F, Brewster DH et al (eds). *Cancer Incidence in Five Continents, Vol. X*. Lyon, IARC, 2013. Available at: <http://ci5.iarc.fr>
9. HAEMACARE Working Group. Manual for coding and reporting haematological malignancies. *Tumori* 2010;96(4):i-A32. Available at: <http://www.tumorionline.it/pdf/HAEMACARE.pdf>
10. Turner JJ, Morton LM, Linet MS et al. InterLymph hierarchical classification of lymphoid neoplasms for epidemiologic research based on the WHO classification (2008): update and future directions. *Blood* 2010;116(20):e90-8.
11. Capocaccia R, De Angelis R. Estimating the completeness of prevalence based on cancer registry data. *Stat Med* 1997;16(4):425-40.
12. Merrill RM, Capocaccia R, Feuer EJ, Mariotto A. Cancer prevalence estimates based on tumour registry data in the Surveillance, Epidemiology, and End Results (SEER) Program. *Int J Epidemiol* 2000;29(2):197-207.
13. ISTAT 2014. Available at: <http://www.demo.istat.it/>
14. Complete Prevalence (ComPrev) Software, Version 2.0 released April 2011. National Cancer Institute. Available at: <http://surveillance.cancer.gov/comprev/>
15. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr* 1974;19(6):716-23.
16. Yu XQ, De Angelis R, Andersson TM, Lambert PC, O'Connell DL, Dickman PW. Estimating the proportion cured of cancer: some practical advice for users. *Cancer Epidemiol* 2013;37(6):836-42.
17. Ederer F, Heise H. *Instructions to IBM 650 programmers in processing survival computations. Methodological note No. 10, End Results Evaluation Section*. National Cancer Institute, Bethesda MD, 1950.
18. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med* 1999;18(4):441-54.
19. Dal Maso L, Guzzinati S, Buzzoni C et al. Long-term survival, prevalence, and cure of cancer: a population-based estimation for 818 902 Italian patients and 26 cancer types. *Ann Oncol* 2014, 25(11): 2251-60.
20. Francisci S, Capocaccia R, Grande E et al. The cure of cancer: a European perspective. *Eur J Cancer* 2009;45(6):1067-79.
21. Lambert PC, Dickman PW, Weston CL, Thompson JR. Estimating the cure fraction in population-based cancer studies by using finite mixture models. *Applied Statistics* 2010;59(1):35-55.
22. Andersson TM, Dickman PW, Eloranta S, Lambert PC. Estimating and modelling cure in population-based cancer studies within the framework of flexible parametric survival models. *BMC Med Res Methodol* 2011;11:96.
23. Othus M, Barlogie B, Leblanc ML, Crowley JJ. Cure models as a useful statistical tool for analyzing survival. *Clin Cancer Res* 2012;18(14):3731-6.
24. Cvancarova M, Aagnes B, Fosså SD, Lambert PC, Møller B, Bray F. Proportion cured models applied to 23 cancer sites in Norway. *Int J Cancer* 2013;132(7):1700-10.
25. Smastuen M, Aagnes B, Johannsen TB, Møller B, Bray F. *Long-term cancer survival: patterns and trends in Norway 1965-2007*. Oslo, Cancer Registry of Norway, 2008. Available at: www.krefregisteret.no/Global/Publikasjoner%20og%20rapporter/CIN2007_del2.pdf
26. Janssen-Heijnen ML, van Steenberghe LN, Voogd AC et al. Small but significant excess mortality compared with the general population for long-term survivors of breast cancer in the Netherlands. *Ann Oncol* 2014;25(1):64-8.
27. Baade PD, Youlden DR, Chambers SK. When do I know I am cured? Using conditional estimates to provide better information about cancer survival prospects. *Med J Aust* 2011;194(2):73-7. Erratum in: *Med J Aust* 2011;194(7):376.
28. Lambert PC, Dickman PW, Osterlund P, Andersson T, Sankila R, Glimelius B. Temporal trends in the proportion cured for cancer of the colon and rectum: a population-based study using data from the Finnish Cancer Registry. *Int J Cancer* 2007;121(9):2052-9.
29. Andersson TM, Lambert PC, Derolf AR et al. Temporal trends in the proportion cured among adults diagnosed with acute myeloid leukaemia in Sweden 1973-2001, a population-based study. *Br J Haematol* 2010;148(6):918-24.
30. Woods LM, Racht B, Lambert PC, Coleman MP. 'Cure' from breast cancer among two populations of women followed for 23 years after diagnosis. *Ann Oncol* 2009;20(8):1331-6.
31. Janssen-Heijnen MLG, Gondos A, Bray F et al. Clinical relevance of conditional survival of cancer patients in Europe: age-specific analyses of 13 cancers. *J Clin Oncol* 2010;28(15):2520-8.
32. Ellison LF, Bryant H, Lockwood G, Shack L. Conditional survival analyses across cancer sites. *Health Rep* 2011;22(2):21-5.
33. Andrae B, Andersson TM, Lambert PC et al. Screening and cervical cancer cure: population based cohort study. *BMJ* 2012;344:e900.
34. Verdecchia A, De Angelis G, Capocaccia R. Estimation and projections of cancer prevalence from cancer registry data. *Stat Med* 2002;21(22):3511-26.
35. Mariotto AB, Yabroff KR, Feuer EJ, De Angelis R, Brown M. Projecting the number of patients with colorectal carcinoma by phases of care in the US: 2000-2020. *Cancer Causes Control* 2006;17(10):1215-26.