

# UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Medicina e Chirurgia

Dipartimento di Medicina del Lavoro “L. Devoto”

Sezione di Statistica Medica e Biometria “G. A. Maccacaro”

Corso di Dottorato in **STATISTICA BIOMEDICA – XXIV CICLO**



TESI DI DOTTORATO DI RICERCA

**“Legame tra l’immunodeficienza HIV-correlata e l’insorgenza di tumori: aspetti di metodologia statistica”**

Dottorando: **Dott. Luigino Dal Maso**

Tutor: **Prof. Adriano Decarli**

Anno Accademico 2010/2011

*Chi ascolta dimentica, chi vede ricorda, chi fa impara*

## **Ringraziamenti**

Sono numerosissime le persone di cui sono debitore per questa tesi e per i risultati del progetto CARL (Cancer and AIDS Linkage Study) e la maggior parte sono citate tra gli autori ed i collaboratori dello studio nelle numerose pubblicazioni, dal 1995 in poi. Mi hanno insegnato moltissimo e di tutti loro ho apprezzato la competenza. La mia gratitudine va anche ai rilevatori e ai collaboratori dei registri, spesso personale precario, raramente ricordati nei lavori scientifici, il cui lavoro è fondamentale per la qualità di questi studi e dei dati dei registri.

Al termine di questo percorso formativo ringrazio i miei genitori che sempre mi hanno incoraggiato nel mio percorso di formazione.

Più di tutti, sono debitore con Claudia Braga che ha sacrificato molto del suo tempo per aiutarmi in questa impresa e alle cui capacità professionali il progetto CARL deve moltissimo.

# INDICE

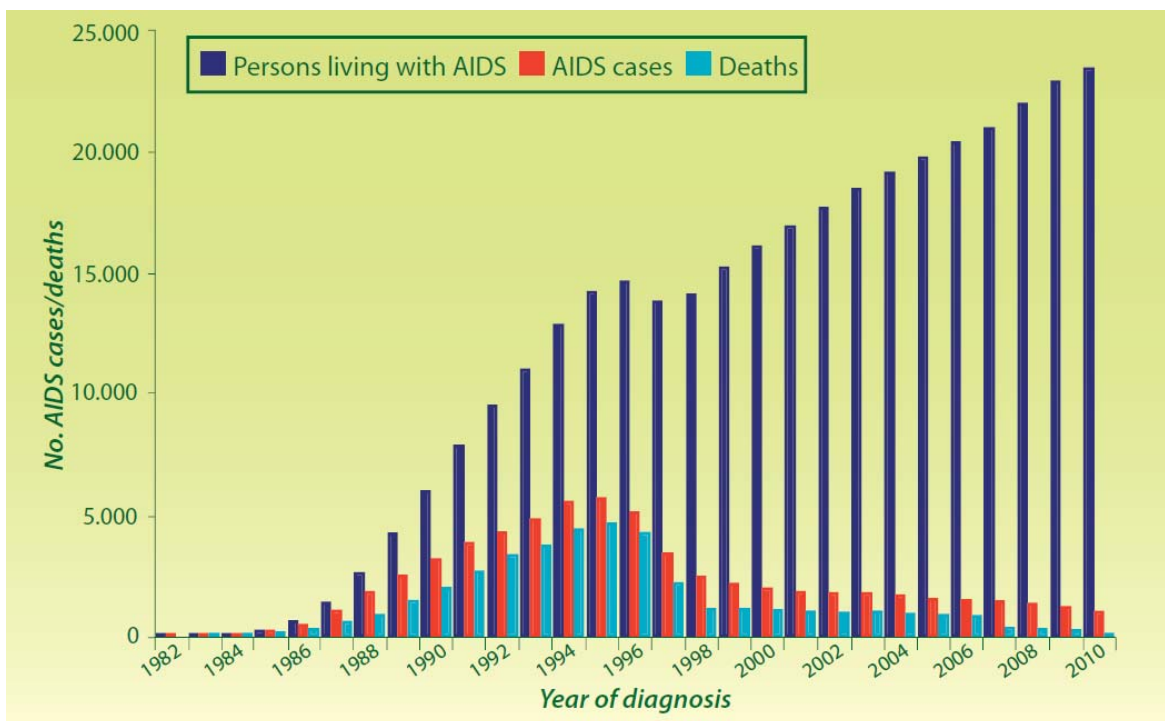
<b>1. INTRODUZIONE E OBIETTIVI</b>	<b>4</b>
<b>2. DESCRIZIONE DEI PROBLEMI METODOLOGICI</b>	<b>5</b>
2.1 Perdite al follow-up .....	5
2.2 Confronti interni e confronti esterni .....	8
<b>3. MATERIALI E METODI</b>	<b>8</b>
3.1 Il registro nazionale AIDS .....	8
3.2 I registri tumori di popolazione.....	9
3.3 La procedura di record-linkage .....	11
3.4 Criteri di inclusione dei casi .....	13
3.5 Disegno dello studio e selezione dei controlli.....	14
3.6 Follow-up.....	19
3.7 Analisi statistiche.....	19
<b>4. RISULTATI DELLO STUDIO</b>	<b>20</b>
4.1 Analisi generali per periodo.....	20
4.2 Analisi stratificate .....	28
4.3 Analisi di sensibilità.....	30
<b>5. DISCUSSIONE</b>	<b>31</b>
<b>6. APPENDICE</b>	<b>34</b>
<b>7. BIBLIOGRAFIA</b>	<b>39</b>

## 1. INTRODUZIONE E OBIETTIVI

All'inizio degli anni '80 si è diffusa a livello mondiale una nuova epidemia dovuta all'infezione da parte del virus dell'immunodeficienza umana (HIV) che induce un progressivo impoverimento delle proprie difese immunitarie<sup>IARC 1996</sup>. La presenza dell'infezione da HIV e di una o più patologie opportunistiche è la condizione necessaria alla definizione della sindrome di immunodeficienza acquisita, meglio conosciuta con l'acronimo inglese AIDS (Acquired Immuno-Deficiency Syndrome), che rappresenta la fase terminale dell'infezione da HIV. Fin dai primi anni dell'epidemia da HIV sono stati segnalati i primi "cluster" di tumori rari, quali i sarcomi di Kaposi<sup>Hymes 1981, IARC 1996</sup>.

La storia naturale dell'infezione da HIV è cambiata drasticamente dopo l'introduzione delle terapie antiretrovirali avvenuta tra il 1995 ed il 1996. Tali terapie, denominate HAART (highly active antiretroviral therapies) o cART (combination antiretroviral therapies), hanno consentito un recupero immunitario notevole (>100%) attraverso un fortissimo contenimento della replicazione virale.

**Figura 1. Casi di AIDS, mortalità AIDS-correlata e numero di casi prevalenti di AIDS in Italia per anno di diagnosi al 31 Dicembre 2010**



Fonte COA 2011

L'aumento dell'attesa di vita dopo la diagnosi di HIV (o dopo l'AIDS) è stato considerevole<sup>Serraino 2009</sup> e continuo dopo il 1995 arrivando, nel periodo 2002-2004, ad una mediana di sopravvivenza di quasi 25 anni dall'infezione da HIV per i pazienti trattati con i più recenti regimi antiretrovirali.<sup>Lima 2007, Harrison 2010</sup> Per quasi la totalità dei pazienti con HIV o AIDS (PHA), l'inizio dei trattamenti induce una fortissima riduzione della carica virale ed un notevole recupero del numero di linfociti CD4+, in modo da rendere abbastanza rara l'insorgenza di patologie opportunistiche. Questo ha portato, in Italia ed in tutti i paesi più ricchi, ad un impressionante calo nel numero di nuovi casi di AIDS, come mostrato in Figura 1. La mortalità ha mostrato andamenti analoghi (Figura 1), con un conseguente aumento continuo dei pazienti viventi con HIV/AIDS.

Un recente studio condotto in Italia<sup>Serraino 2009</sup> ha mostrato che circa 2/3 dei PHA sono vivi 5 anni dopo la diagnosi di AIDS, con un sostanziale miglioramento della sopravvivenza dopo l'AIDS rispetto al periodo pre-cART (prima del 1995) in cui era circa la metà.

Gli aspetti di metodologia statistica relativi alle analisi di incidenza ed eccesso di rischio sono state discusse nella relazione del secondo anno del presente dottorato di ricerca in Statistica Biomedica.<sup>Dal Maso 2009, Polesel 2010</sup>

Nel corso del terzo anno di dottorato sono stati affrontati i problemi che emergono negli studi di sopravvivenza dopo un tumore nei PHA in Italia. Saranno descritte le analisi eseguite per uno studio che documenta la sopravvivenza dopo i tumori nei PHA durante tutto il corso dell'epidemia (1986-2005) e la confronta con quella nei pazienti con tumore HIV-negativi. Sono state effettuate analisi stratificate e alcune simulazioni che hanno permesso di approfondire ed esaminare in dettaglio limiti e potenzialità dei diversi disegni di studio e analisi statistiche condotte.

## **2. DESCRIZIONE DEI PROBLEMI METODOLOGICI**

### **2.1 Perdite al follow-up**

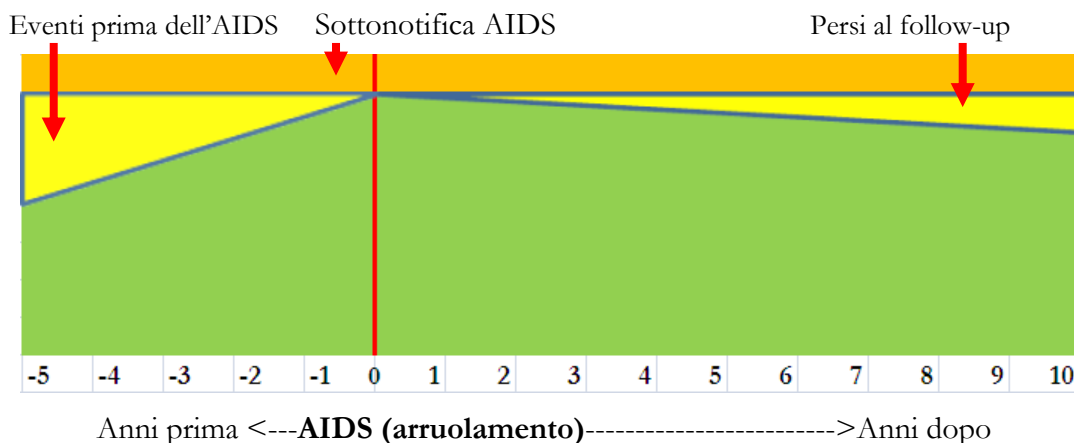
La validità di uno studio prospettico dipende fondamentalmente dalla completezza della registrazione degli eventi di interesse (tumori, morti) e dalla corretta computazione della popolazione a rischio.

La Figura 2 sintetizza i possibili bias legati all'uso delle coorti prospettiche definite dalla diagnosi di AIDS al fine di stimare l'incidenza o la sopravvivenza dopo il tumore nei PHA.

In sintesi, nelle stime di sopravvivenza e rischio di morte in PHA con tumore, possono verificarsi due principali tipi di distorsioni:

1. Perdite al follow-up diverse nei PHA rispetto alla popolazione generale;
2. Mortalità per tumore nei pazienti con HIV che induce una sottotifica differenziale di AIDS (i pazienti muoiono senza che l'AIDS venga registrato).

**Figura 2. Possibili bias legati all'uso delle coorti prospettive definite dalla diagnosi di AIDS**



*In Verde il periodo di osservazione nelle coorti AIDS*

*In Giallo le perdite al follow-up sia prima che dopo l'AIDS*

*In Arancio la possibile sottotifica dei casi di AIDS*

Il primo bias è relativamente contenuto a due condizioni:

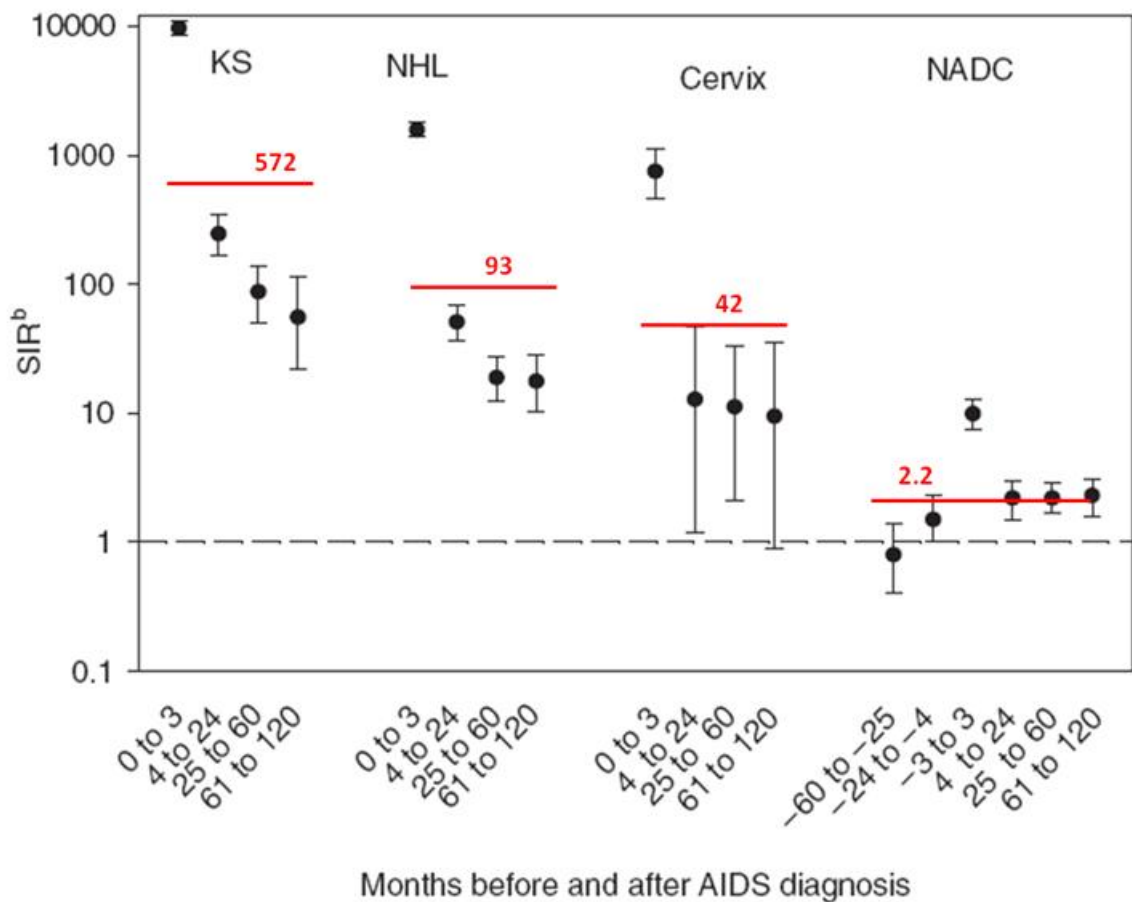
- a) che il registro effettui un follow-up attivo dello stato in vita dei soggetti, o
- b) che gli studi siano condotti attraverso procedure di record-linkage affidabili che incrociano i dati delle coorti AIDS e quelli dei registri tumori con registri di mortalità in modo da garantire un follow-up pressoché completo.

Il secondo bias si verifica in conseguenza del fatto che i pazienti HIV-positivi con tumore che si verifica prima della diagnosi di AIDS, per essere inclusi nello studio, devono sopravvivere fino all'AIDS stesso. Di conseguenza, è probabile che una quota non trascurabile di casi (in particolare i tumori rapidamente fatali) non sia registrata. La quantificazione e l'eventuale rimozione del secondo bias è più difficile.

Negli studi di sopravvivenza, il follow-up deve necessariamente partire dalla data dell'evento (incidenza di tumore, in questo caso). Negli studi di questo tipo in PHA rimane da definire la "finestra di osservazione" rispetto alla diagnosi di AIDS. Non è possibile utilizzare la finestra utilizzata negli studi di incidenza ed eccesso di rischio<sup>Dal Maso 2009, Polesel 2010</sup>, che

includeva i casi di tumore da 5 anni prima dell'AIDS, in quanto la quota di pazienti con HIV che sviluppa un tumore e muore prima di sviluppare l'AIDS <sup>Melbye 1994, Grulich 1999, Li 2002</sup> è tanto più elevata quanto più la finestra si estende prima dell'AIDS. D'altro canto, l'inclusione solo dei casi di tumore nel periodo successivo all'AIDS <sup>Engels 2006</sup>, eliminando la maggior parte degli eventi, consente di studiare solamente quelli che si verificano nella fase più tardiva della storia naturale dell'infezione da HIV.

**Figura 3. Rischio di tumore (Standardized Incidence Ratio-SIR) nei pazienti con AIDS al variare del tempo rispetto alla diagnosi di AIDS in Italia, 1997-2004**



Fonte Dal Maso et al 2009

È fondamentale ricordare che l'AIDS è una sindrome <sup>Ancelle-Park 1993, COA 2011</sup> e, come tale, la diagnosi è molto sensibile al variare dell'intensità diagnostica che può essere modificata dall'insorgenza di altre serie patologiche, quali i tumori. I risultati di diversi studi <sup>Frisch 2001</sup>, compresi quelli condotti negli anni passati del presente dottorato, <sup>Dal Maso 2009</sup> hanno mostrato che si verifica un'importante anticipazione diagnostica dei tumori in concomitanza della diagnosi di AIDS e, allo stesso tempo, dell'AIDS in caso di diagnosi di tumore (Figura 3).

Infine, il cambiamento del significato clinico della diagnosi di AIDS nel corso dei decenni e dei Paesi <sup>Kaldor 2008</sup> rende difficilmente confrontabili nel tempo e nello spazio le sopravvivenze (così come le incidenze) dei tumori dopo l'AIDS.

## 2.2 Confronti interni e confronti esterni

I problemi descritti nel paragrafo 2.1 suggeriscono alcune cautele nell'esaminare (**confronto "interno"**) le differenze temporali nelle stime di sopravvivenza ai PHA. <sup>Hessol 2007</sup> Dal punto di vista clinico, una domanda rilevante riguarda la differenza di sopravvivenza tra i pazienti con HIV e tumori rispetto alla sopravvivenza dei pazienti con tumore ma senza HIV (**confronto "esterno"**) (popolazione generale). Tale quesito (confronto "esterno") è di rilevanza crescente in considerazione dell'aumentata sopravvivenza dei pazienti con AIDS. <sup>Serraino 2009</sup>

Nel confronto di sopravvivenze in PHA e popolazione generale va tenuto conto, in primo luogo, della diversa distribuzione per sesso ed età dei tumori nei PHA rispetto ai pazienti senza AIDS.

Numerosi studi hanno analizzato la sopravvivenza solo per alcune sedi tumorali in PHA <sup>Dal Maso 2001b, Brau 2007; Oehler-Janne 2008; Chao 2010; Lodi 2010</sup> oppure confrontandola a quelle nei pazienti HIV-negativi arruolati in serie cliniche <sup>Puoti 2004, Spagnuolo 2011</sup>.

L'unico studio che ha effettuato un confronto su base di popolazione, tenendo conto che i tumori in PHA hanno una distribuzione per sesso ed età molto diversa dai pazienti con tumore senza HIV, è stato condotto nello stato di New York, Stati Uniti <sup>Biggar 2005</sup>. Con il principale obiettivo di valutare il cambiamento nel tempo della sopravvivenza dopo il tumore dei PHA e di confrontare il rischio di morte nei PWA rispetto ai pazienti senza AIDS, lo studio stimava la sopravvivenza dopo 24 mesi dal tumore ed il rischio di morte (Death Hazards) aggiustato per sesso, razza, età e periodo di calendario. Il gruppo di confronto utilizzato includeva, quindi, tutti i pazienti con tumore senza AIDS. <sup>Biggar 2005</sup>

## 3. MATERIALI E METODI

### 3.1 Il registro nazionale AIDS

Nelle analisi statistiche descritte in questa relazione sono stati utilizzati i dati del Registro Nazionale AIDS aggiornati al 31 dicembre 2005. La notifica dei casi di AIDS al Registro nazionale AIDS è obbligatoria in tutta Italia dal novembre 1986. Alla fine del 2005, un totale di 57.531 casi di AIDS sono stati notificati, <sup>COA 2008</sup> di cui il 32% dopo l'introduzione delle



terapie antiretrovirali (nel 1996). Del totale dei casi diagnosticati il 78% erano di sesso maschile, l'11,3% in età pediatrica. In totale il 62% dei pazienti risultavano deceduti.

La completezza e la qualità del Registro Nazionale AIDS sono soddisfacenti. Uno studio che ha linkato il Registro AIDS al Registro ISTAT dei certificati di morte effettuato nel 1992 ha stimato che la sottostima dei casi di AIDS in Italia (Figura 2, rettangolo rosso in alto) era inferiore al 10% <sup>Conti 1992</sup>, una delle percentuali più basse di sottotifica registrate in Europa <sup>Ajdacic-Gross 2001</sup>.

L'età mediana alla diagnosi dei casi adulti, sia maschi che femmine, mostra un aumento nel tempo. L'andamento dei tassi d'incidenza per anno di diagnosi mostra un incremento dei casi fino al 1995, seguito da una diminuzione nel 1996 continuata fino al 2001, mentre dal 2002 il numero dei casi diagnosticati sembra stabilizzarsi. Inoltre, si nota un aumento della proporzione dei casi attribuibili alla trasmissione sessuale e una diminuzione di quella delle altre modalità di trasmissione. Solo il 35% dei malati ha fatto uso di terapie antiretrovirali. <sup>COA2008</sup>

È importante osservare che la proporzione di pazienti che non sapevano di essere sieropositivi alla diagnosi di AIDS è molto aumentata dal 1996 (21%) al 2006 (oltre la metà dei pazienti) ed è più elevata tra coloro che hanno acquisito l'infezione attraverso rapporti sessuali e tra gli stranieri, arrivano cioè allo stadio di AIDS conclamato ignorando di essere infetti. Tale osservazione è importante dal punto di vista clinico in quanto il quadro immunologico e virologico al momento della diagnosi di AIDS in pazienti che hanno effettuato almeno un precedente trattamento antiretrovirale è certamente migliore rispetto a coloro che non hanno effettuato alcuna terapia. <sup>COA2008</sup> Anche il quadro delle patologie di esordio clinico è differente tra trattati e non trattati. In particolare, si evidenzia tra i non trattati una proporzione più elevata di candidosi, linfomi, polmoniti ricorrenti e carcinoma cervicale invasivo. Di tali differenze si deve tener conto anche quando si stimano le incidenze dei tumori e le loro variazioni nel tempo.

Ogni record nominativo del Registro AIDS comprende i dati anagrafici della persona, nonché informazioni sulla data di diagnosi di AIDS con relativa patologia e data di notifica. La presenza di informazioni sul comune di nascita e di residenza o, l'eventuale, data di morte ha consentito di poter verificare l'esattezza degli accoppiamenti generati dalla procedura di linkare (Paragrafo 3.3).

### **3.2 I registri tumori di popolazione**

L'Associazione Italiana dei Registri Tumori (AIRTUM), attiva da oltre 20 anni, comprendeva alla fine degli anni '90, ben 24 Registri indipendenti che includevano una

popolazione di circa 17 milioni di persone (Tabella 1), pari al 30% del totale dei residenti in Italia. AIRT 2006, Dal Maso 2009

I Registri Tumori comprendevano le regioni Friuli Venezia Giulia, Romagna, Umbria, e parte del Veneto, i comuni di Milano e Torino, le province di Bolzano (Alto Adige), Biella, Ferrara, Firenze e Prato, Genova, Macerata, Modena, Parma, Ragusa, Reggio Emilia, Salerno, Sassari, Sondrio, Siracusa, Trento, Varese, e parte delle province di Brescia e Napoli. I Registri Tumori variano considerevolmente per la popolazione inclusa. Si passa da meno di 200.000 casi per Sondrio a oltre due milioni per il Veneto; anche il numero di anni di registrazione è estremamente variabile (Tabella 1).

Nel complesso i dati raccolti dai Registri Tumori italiani hanno mostrato una buona qualità globale sia in termini di completezza che di validità dei casi raccolti. AIRT 2006, Curado 2008

**Tabella 1. Periodo e popolazione coperta e casi di AIDS nel 1986-2005 nelle aree coperte dai 24 Registri Tumori Italiani**

<b>Registri Tumori</b>	<b>Periodo di osservazione</b>	<b>Popolazione (x1.000)<sup>a</sup></b>	<b>Casi di AIDS</b>
Alto Adige	1995-2002	460	230
Biella	1994-2002	189	268
Brescia	1996-2001	1.012	1.909
Ferrara	1986-2005	314	959
Firenze e Prato	1986-2003	1.162	1.199
Friuli Venezia Giulia	1995-2003	1.188	407
Genova	1986-2003	920	1.775
Macerata	1991-2000	293	135
Milano	1996-2005	1.256	4.822
Modena	1988-2004	615	633
Napoli	1996-2003	541	110
Parma	1986-2003	394	368
Ragusa	1981-2003	291	63
Reggio Emilia	1996-2005	450	412
Romagna	1986-2005	803	1.948
Salerno	1996-2001	1.088	223
Sassari	1992-2002	469	374
Sondrio	1998-2002	177	118
Siracusa	1998-2003	396	157
Trento	1995-2005	460	347
Torino	1987-2003	1.091	1.794
Umbria	1994-2003	831	435
Varese	1986-2002	800	1.668
Veneto	1987-2002	2.077	1.599
<b>Totale</b>		<b>17.277</b>	<b>21.951</b>

<sup>a</sup>Nel periodo 1997-2002.

### 3.3 La procedura di record-linkage

Date le dimensioni spesso ridotte degli archivi dei Registri Tumori (minori, salvo rare eccezioni, di 100.000 records), l'interesse si è rivolto verso lo sviluppo di un applicativo di *linkage* semiautomatico che avesse come punto di forza un'alta sensibilità (prossima al 100%) in modo da limitare quanto più possibile la perdita di possibili accoppiamenti.

Un'altra priorità risiedeva nello sviluppo di un applicativo in grado di tener conto dei più comuni errori nell'inquadratura di nomi italiani, quali la trascrizione di uno solo dei costituenti del nome, o cognome, nel caso di nomi, o cognomi, doppi o tripli (es. Levi al posto di Levi Montalcini o Maria invece di Maria Teresa), gli errori nei nomi contenenti apostrofi (D'Avanzo con Davanzo o Da Avanzo) o nell'uso delle vocali accentate (Faé invece di Fae o Faè).

La realizzazione di un applicativo specifico per la lingua italiana, e per tutte le lingue d'origine latina vista l'implementazione del controllo sulle lettere speciali come la ç o la ñ, colma la lacuna della non disponibilità di software specifici per idiomi diversi da quello anglosassone.

Sono stati quindi utilizzati tre soli campi chiave per l'identificazione dei soggetti ovvero il cognome, il nome e la data di nascita, vista la loro rilevanza e vista la loro obbligatoria presenza nel Registro AIDS e in tutti i Registri Tumori.

Nell'organizzazione dello studio è stata esplicitamente richiesta l'obbligatoria osservanza della tutela della privacy dei soggetti inclusi nei due tipi di Registri. Questa richiedeva lo svolgimento delle operazioni di linkage direttamente nelle sedi dei diversi Registri Tumori Italiani, utilizzando una copia del Registro Nazionale AIDS, al fine di estrarre informazioni in modo anonimo. Era quindi necessario sviluppare di un applicativo estremamente trasportabile, con minime richieste hardware, in modo da poter prescindere dalle caratteristiche dei computer contenenti i dati del Registro in cui fisicamente si sarebbero svolte le operazioni di linkage.

La scelta del linguaggio di programmazione è caduta sul linguaggio Clipper per la gestione di file di dati in formato dBase (estensione DBF). Il vantaggio della scelta del formato DBF risiede nell'esistenza di semplici strumenti di conversione al formato dBase in tutti i software normalmente utilizzati per la gestione di basi di dati come Access o Oracle e/o di analisi statistica SAS o SPSS, utilizzati per la realizzazione dei Registri.

La scelta del linguaggio Clipper fornisce inoltre alcuni importanti vantaggi, il più importante è il fatto di essere stato specificatamente sviluppato per la gestione di files in formato dBase.

L'utilizzo infine di un applicativo con l'interfaccia grafica DOS che non necessita di procedure di installazione, ha consentito e consente tuttora l'impiego del software anche su macchine con una ridotta dotazione hardware.

Con queste caratteristiche di base è stato quindi realizzato appositamente un applicativo, denominato SALI acronimo per "Software for Automatic Linkage in Italy"<sup>Dal Maso 2001a</sup>. In seguito, il software è stato sviluppato (Tabella 2)<sup>Dal Maso 2010a</sup> in modo da aggiornare le procedure, consentendo di "linkare" campi di lunghezza maggiore (doppi nomi o cognomi, comuni in Svizzera nelle donne che, sposandosi, acquisiscono il cognome del marito)<sup>Clifford 2005</sup> e verificare anche l'inversione dei campi nome e cognome, evento possibile per gli stranieri il cui numero non è trascurabile nei PHA in Italia negli anni più recenti.

**Tabella 2. Condizioni per l'accoppiamento dei records nelle diverse fasi del programma SALI, versione 3.3** *Fonte Dal Maso et al 2010a*

Fase <sup>a</sup>	Key variables			User intervention required
	<i>Surname</i>	<i>Name</i>	<i>Date of birth</i>	
0	Equal	One of the two names contained in the other	Equal	No
1 <sup>b</sup>	As above	As above <i>or</i> first 7 characters equal <i>or</i> one name without the first letter contained in the other	Year equal	No
2	As above	As above	At least 7 common digits <i>and</i> at least 5 digits in the same position	Yes
3	For each surname a maximum of 20 characters not found in the other <i>and</i> a common substring at least 2 characters long	As above	Equal (month and day allowed to be reversed)	Yes
4	One of the two surnames contained in the other <i>or</i> first 7 characters equal	For each name a maximum of 20 characters not found in the other <i>and</i> a common substring at least 2 characters long	Equal (month and day allowed to be reversed)	Yes
5	One of the two surnames contained in the other <i>or</i> first 7 characters equal	Equal	At least 7 common digits <i>and</i> at least 6 digits in the same position	Yes
6 <sup>c</sup>	For each surname a maximum of 3 characters not found in the other <i>and</i> a common substring at least 4 characters long	One of the two names is equal to the first part of the other	At least 7 common digits <i>and</i> at least 6 digits in the same position	Yes

<sup>a</sup> Matches already performed in a previous stage are skipped (Stages may be skipped).

<sup>b</sup> Stage 1 performed after Stage 5.

<sup>c</sup> Stage 6 can be performed if necessary (not in current study).

Il software opera l'incrocio dei record (record linkare) utilizzando due files di input, nel formato DBF, che includano almeno i seguenti campi chiave: Nome, Cognome, Data di nascita (separata in 3 campi di formato testo, corrispondenti al giorno, al mese e all'anno di

nascita). La Tabella 2 sintetizza le condizioni richieste perché si verifichi l'accoppiamento tra i record dei due Registri ed il dettaglio del funzionamento è descritto in un recente manoscritto <sup>Dal Maso 2010a</sup>. Per tutte le fasi, un'ispezione manuale di altri campi comuni eventualmente presenti nei due files di input, permette l'eliminazione di accoppiamenti non corretti.

Durante lo svolgimento della procedura, il contenuto dei campi Nome e Cognome non è mai visibile agli operatori e, nel file di output, è sostituito da un asterisco come ulteriore garanzia per la salvaguardia della privacy dei pazienti inseriti in entrambe le registrazioni.

La procedura di linkage adottata dal programma SALI ha permesso di raggiungere una sensibilità praticamente del 100%. Comparando i risultati ottenuti con SALI con quelli ottenuti grazie a due operazioni di linkage manuale effettuate precedentemente nella provincia di Genova ed in quella di Firenze, sui dati dei locali Registri Tumori e del Registro Nazionale AIDS, è stato possibile dimostrare una perfetta concordanza nell'identità dei soggetti linkati (44 nella provincia di Genova e 25 in quella di Firenze) <sup>Franceschi 1998</sup>. La stessa validazione è stata effettuata in Svizzera attraverso uno studio ad hoc condotto nel Registro di Ginevra dove è emersa una sensibilità del 99% (441 su 445). <sup>Clifford 2005</sup>

La specificità del programma non supera comunque il 70%, richiedendo quindi una parte di ispezione manuale da parte dell'operatore, per scartare, in base ad informazioni addizionali disponibili nelle due registrazioni, come la data di morte o il luogo di nascita e/o residenza, gli accoppiamenti non corretti.

### **3.4 Criteri di inclusione dei casi**

I casi di tumore in PHA inclusi inizialmente nello studio soddisfacevano le seguenti condizioni:

- (i) erano stati diagnosticati con AIDS tra il 1986 e il 2005;
- (ii) erano residenti in un'area in cui è attivo un Registro Tumori;
- (iii) avevano un'età compresa tra 16 e 74 anni alla diagnosi di tumore;
- (iv) avevano una data di tumore non oltre 12 mesi prima della diagnosi di AIDS.

Le diagnosi tumorali sono state ricodificate secondo la decima revisione della Classificazione Internazionale delle Malattie (ICD-10)<sup>WHO 1992</sup> e sono state verificate dai coordinatori dei Registri Tumori per quanto riguarda tipo istologico e sede. La base di diagnosi è stata categorizzata in casi con conferma microscopica (istologica, ematologica o citologica) o con altri tipi di modalità diagnostica (clinica, strumentale o attraverso i certificati di morte che rappresentavano il 6,5% dei casi nel periodo pre-HAART e meno dell'1% nel periodo HAART).

Nei casi in cui un tumore definitorio di AIDS fosse registrato sia al Registro Nazionale AIDS che ad un Registro Tumori, la data di diagnosi del tumore è stata definita come la prima fra le due. Nei casi in cui un tumore definitorio di AIDS fosse stato notificato ad un Registro Tumori prima della data di diagnosi di AIDS riportata al Registro Nazionale AIDS, la data di AIDS è stata anticipata.

Nelle analisi di sopravvivenza, a differenza da quelle utilizzate per le stime di incidenza e di rischio, sono state incluse anche le diagnosi di tumore (Sarcomi di Kaposi, linfomi non-Hodgkin e carcinomi invasivi della cervice) notificate al RAIDS ma non in un Registro Tumori. In questo caso, la morfologia dei tumori è stata considerata “ignota” (NOS - not otherwise specified).

### 3.5 Disegno dello studio e selezione dei controlli

Per confrontare la sopravvivenza nei PHA con tumore rispetto a quella nei pazienti con tumore senza AIDS sono possibili diversi disegni di studio, i principali sono:

1. **Coorte:** la popolazione viene definita sulla base di una esperienza “iniziale” comune a tutti gli individui che la compongono (tumore)<sup>Breslow e Day 1987</sup> e i PHA ne rappresentano un sottoinsieme.

2. **Caso-coorte:** a partire dalla intera coorte originale viene individuata, attraverso un procedimento di selezione casuale, una sotto-coorte costituita da una percentuale definita di soggetti.<sup>Breslow 2009</sup> Ogni individuo della coorte avrà la stessa probabilità di entrare a far parte della sotto-coorte.

3. **Caso-controllo annidato (nested):** i casi sono rappresentati dai pazienti con una certa caratteristica (tumore ed AIDS, nel nostro caso). Dall'intera coorte si selezionano uno o più controlli che non hanno la caratteristica di interesse (nel nostro caso i tumori senza AIDS) al momento in cui l'ha sviluppata il caso. Possono essere appaiati ai casi per una serie di caratteristiche ritenute possibili confondenti dell'*outcome* in studio.

In pratica, la scelta del disegno dello studio è spesso dettata da considerazioni contingenti quali i costi o la disponibilità di accesso alle intere coorti. L'accesso completo alla banca dati individuale anonima AIRTUM, che include oltre 1.300.000 records, risulta condizionato a una richiesta formale ai singoli Registri (proprietari dei dati) e viene raramente concesso al di fuori di progetti istituzionali dell'AIRTUM (studi di incidenza<sup>AIRTUM 2006</sup>, sopravvivenza<sup>AIRTUM 2011</sup> e prevalenza<sup>AIRTUM 2010</sup>).

La necessità di utilizzare un sottoinsieme dell'intera banca dati da cui estrarre i controlli, ha suggerito l'utilizzo di un disegno di studio di tipo “caso-controllo nested” ed è stata richiesta la

parte di bancadati anonima (circa 500.000 record) contenente tutti i casi di tumore per le categorie per la cui sede, sesso e classe di età era presente almeno un tumore nelle PHA.

Per ciascun caso di tumore in PHA, il numero di controlli da estrarre è stato definito tenendo conto della frequenza dei casi di tumore nei PHA e nell'intera bancadati per sede, sesso ed età alla diagnosi di tumore. Per alcuni tipi di tumore (linfomi, tumori della cervice uterina e della pelle), è stata anche considerata la frequenza per sede di insorgenza (es.: linfomi primitivi cerebrali) e il sottotipo istologico (es.: linfomi di Burkitt o tumori squamocellulari della cervice uterina).

L'ispezione delle frequenze nei PHA e nella "sottocoorte" AIRTUM, ha suggerito di effettuare un accoppiamento 1:1 per i sarcomi di Kaposi (raro nei giovani senza AIDS), di 1:2 per i linfomi non-Hodgkin e 1:5 per tutte le altre sedi e sottosedie tumorali.

Sono stati considerati "non-AIDS" i pazienti con tumore notificati presso la bancadati AIRTUM che non avevano, tra i PHA, corrispondenti pazienti con lo stesso tumore, stesso sesso, data di nascita e data di diagnosi di tumore.

L'accoppiamento (matching) è stato effettuato per sede (prime 2 cifre ICD10), istologia (per alcune sedi), sesso, età alla diagnosi di tumore (in quinquenni), periodo di diagnosi (1986-1991, 1991-1995, 1996-2000, 2001-2005) e area di residenza (Nord, Centro e Sud).

La procedura di estrazione automatica dei controlli è stata effettuata utilizzando un programma SAS che utilizza alcune macro descritte in dettaglio nell'Appendice 1. In breve, è possibile specificare il numero di controlli da estrarre casualmente per ciascun caso ed il numero (massimo 6) delle variabili di stratificazione. Per ciascuna variabile di stratificazione è possibile estrarre controlli in una cella adiacente (massima distanza 1) a quella del caso (es. caso di età 40-44, controllo 45-49).

L'output della Macro consiste in due datasets (Appendice 1):

1. uno con tutti i record dei casi e dei controlli estratti che contiene tutte le variabili del dataset di input e, sia per i casi che per i controlli, l'identificativo (`_idMatch`) del caso per il quale sono stati estratti i controlli, in modo da consentire successive analisi "individual matched";

2. l'altro con tutti i controlli non matchati, utili per l'integrazione manuale di controlli qualora non sia stato raggiunto il numero di controlli richiesto.

Solo per le variabili di stratificazione età, periodo di diagnosi e area di residenza (non per sede e sesso), il programma è stato impostato in modo da estrarre alcuni controlli nelle celle adiacenti a quella del caso, qualora non fossero stati individuati gli n controlli richiesti con le stesse caratteristiche.

La Tabella 3 mostra la frequenza dei tumori definitivi di AIDS (sarcomi di Kaposi, linfomi non-Hodgkin e tumori invasivi della cervice uterina) nei PHA (casi) e nelle persone senza HIV (controlli) verificatisi nelle aree coperte da Registro Tumori tra il 1986 e il 2005. Sono stati esclusi i casi di sarcoma di Kaposi (941) per i quali non è stato possibile estrarre dei controlli a causa della diversa distribuzione per età rispetto alla popolazione HIV-negativa. <sup>Dal</sup>

Maso 2005

**Tabella 3. AIDS-defining cancers in persons with AIDS (PHA)<sup>1</sup> and matched controls<sup>2</sup> by year of cancer diagnosis in Italy, 1986-2005**

ICD10; Cancer type or site	1986-2005		1986-1995		1996-2005	
	AIDS		AIDS		AIDS	
	Yes	No	Yes	No	Yes	No
<b>C46; Kaposi sarcoma<sup>3</sup></b>	<b>585</b>	<b>585</b>	<b>329</b>	<b>167</b>	<b>256</b>	<b>418</b>
<b>C82-C85,C88,C96; Non-Hodgkin Lymphomas<sup>4</sup></b>	<b>1326</b>	<b>2652</b>	<b>636</b>	<b>1010</b>	<b>690</b>	<b>1642</b>
<i>CNS (Topography 70-72, all histologies)</i>	73	146	38	42	35	104
<i>DLBC</i>	202	404	51	101	151	303
<i>DLBC Immunoblastic</i>	87	174	46	82	41	92
<i>Burkitt</i>	46	92	18	28	28	64
<i>SLL/CLL/Immunoproliferative Diseases</i>	11	22	7	14	4	8
<i>Follicular</i>	11	22	3	6	8	16
<i>Other specified NHL</i>	36	72	17	34	19	38
<i>Unspecified NHL</i>	860	1720	456	703	404	1017
<b>C53; Cervix uteri<sup>5</sup></b>	<b>92</b>	<b>460</b>	<b>27</b>	<b>118</b>	<b>65</b>	<b>342</b>
<i>Squamous</i>	37	185	6	30	31	155
<i>NOS (1 Adenocarcinoma in PHA)</i>	55	275	21	88	34	187

NOS=Not otherwise specified; CNS=Central nervous system ; DLBC=Diffuse large B-cell; SLL/CLL=Small lymphocytic lymphoma/Chronic lymphocytic leukemia

<sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years, between 1986 and 2005 after AIDS diagnosis.

<sup>2</sup>Matched for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

<sup>3</sup>1:1 matching, KS cases in PHA without controls in the same age and sex strata (941) have been excluded

<sup>4</sup>1:2 matching

<sup>5</sup>1:5 matching



**Tabella 4. Non-AIDS-defining cancers in persons with AIDS<sup>1</sup> and matched controls<sup>2</sup> by year of cancer diagnosis in Italy**

ICD10; Cancer type or site	1986-2005		1986-1995		1996-2005	
	AIDS		AIDS		AIDS	
	Yes	No	Yes	No	Yes	No
<b>All Non-AIDS-defining cancers</b>	<b>354</b>	<b>1770</b>	<b>105</b>	<b>507</b>	<b>249</b>	<b>1263</b>
C00-C14, C30-C32; Head and neck	15	75	3	13	12	62
C00-Lip	1	5	1	5	0	0
C01-02 Tongue	1	5	0	0	1	5
C06 Mouth NOS	1	5	0	0	1	5
C09 Tonsil	1	5	0	0	1	5
C10 Oropharynx	1	5	0	0	1	5
C13 Hypopharynx	2	10	0	0	2	10
C30; Nose ear	2	10	1	3	1	7
C31; Acc sinus	1	5	0	3	1	2
C32; Larynx	5	25	1	2	4	23
C15; Oesophagus	2	10	0	0	2	10
C16; Stomach	12	60	3	15	9	45
C17; Small Intestine	1	5	0	0	1	5
C18-C20; Colorectum	23	115	7	35	16	80
C18; Colon	11	55	2	10	9	45
C19; Rectosigmoid junction	4	20	2	10	2	10
C20; Rectum	8	40	3	15	5	25
C21; Anus	14	70	3	15	11	55
C22; Liver	18	90	3	14	15	76
C23-C24; Gallbladder & biliary tract	2	10	0	0	2	10
C25; Pancreas	5	25	2	10	3	15
C26; Dig NS	2	10	0	0	2	10
C33-C34; Trachea and lung	62	310	12	60	50	250
C37-C38; Thy, heart, medias, pleura	1	5	0	0	1	5
C40-C41; Bone and articular cartilages	2	10	1	5	1	5
C43; Melanoma	5	25	3	15	2	10
C44; Skin, non melanoma	39	195	9	44	30	151
Squamous	14	70	1	5	13	65
Basal	18	90	6	30	12	60
NOS	7	35	2	9	5	26
C45; Mesothelioma	1	5	0	0	1	5
C47; Periph. nerves	2	10	0	2	2	8
C48; Peritoneum	1	5	0	0	1	5
C49; Soft/connective tissues	1	5	0	0	1	5
C50; Female breast	5	25	1	5	4	20
C51,C52; Vulva	3	15	0	0	3	15
C54; Endometrium	1	5	0	0	1	5
C55; Uterus NOS	1	5	1	4	0	1
C56; Ovary	1	5	1	5	0	0
C57; Other female genital organs	2	10	0	1	2	9
C60, C63; Penis and NOS male genital organs	3	15	0	2	3	13
C61; Prostate	3	15	1	5	2	10
C62; Testis	4	20	3	15	1	5
C64-C65; Kidney	3	15	1	5	2	10
C66,C68; Ureter and NOS urinary system	1	5	0	0	1	5
C67; Bladder	3	15	2	10	1	5
C68; Urinary NOS	2	10	1	3	1	7
C70-C72; Brain and central nervous system	16	80	6	30	10	50
C76-80; Site NOS	9	45	3	15	6	30
C81; Hodgkin lymphoma	71	355	30	136	41	219
Mixed cell	30	150	13	62	17	88
Nodular sclerosis	16	80	7	35	9	45
Lymphocyte depletion	4	20	4	9	0	11
NOS	21	105	6	30	15	75
C90; Multiple myeloma/plasma cell neoplasms	5	25	3	13	2	12
C91-C95; Leukaemias, all	13	65	6	30	7	35
C91; Lymphoid leukaemia	6	30	3	15	3	15
C92-C94; Myeloid and other spec. leukemia	7	35	3	15	4	20

NOS=Not otherwise specified <sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years, between 1986 and 2005 from 1 year prior AIDS diagnosis (at/after AIDS for AIDS-defining cancers). <sup>2</sup>Matched (1:5) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

La Tabella 4 mostra la distribuzione per sede tumorale e periodo di incidenza dei 354 tumori non-definitivi di AIDS e dei rispettivi 1770 (1:5) controlli selezionati per le analisi relative allo studio. Solo le sedi con almeno 10 casi nell'intero periodo di studio saranno analizzate in dettaglio nelle successive tabelle e figure.

La distribuzione per sesso, età ed anno di diagnosi dei tumori che sono stati utilizzati nelle successive analisi di sopravvivenza è mostrata nella Tabella 5.

**Tabella 5. Cancers in persons with AIDS<sup>1</sup> (Ca) and matched controls<sup>2</sup> (Co) by selected characteristics in Italy, 1986-2005<sup>3</sup>**

ICD10; Cancer type or site	Sex		Age	Year of diagnosis		
	Men Ca:Co	Women Ca:Co	Median Ca:Co	1986-1995 Ca:Co	1996-2005 Ca:Co	Median Ca:Co
<b><i>AIDS-defining cancers</i></b>						
C46; Kaposi sarcoma	516:506	56:53	48:48	320:312	252:247	1995:1999
C82-C85,C88,C96; NHL	1005:1965	245:484	36:39	572:1115	678:1334	1996:1997
C53; Cervix uteri	-	91:452	35:35	26:129	65:323	1998:1998
<b><i>Non-AIDS-defining cancers</i></b>						
C00-C14, C30-C32; Head and neck	5:25	4:20	39:42	2:10	7:35	2000:2000
C16; Stomach	10:50	1:5	41:43	3:15	8:40	1998:1999
C18-C20; Colorectum	18:90	2:10	46:46	6:29	14:70	2001:2001
C21; Anus	10:49	3:15	42:43	3:15	10:49	2000:2000
C22; Liver	13:63	2:9	43:44	2:9	13:63	2000:1999
C33-C34; Trachea and lung	46:227	8:40	46:47	8:39	46:228	1999: 1999
C44; Skin, non melanoma	25:125	7:35	47:47	4:20	28:140	2000: 1999
C70-C72; Brain, central nerv. sys.	7:35	3:14	36:40	4:19	6:30	1997: 1997
C81; Hodgkin lymphoma	50:246	5:25	36:36	19:94	36:177	1999:1998
C91-C95; Leukaemias, all	7:35	2:10	37:37	3:15	6:30	1996: 1998
Other NADC	41:203	18:88	40:40	16:78	43:123	1999:1999

Ca:Co= Case:Controls; NADC=Non-AIDS-defining cancers. <sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years, between 1986 and 2005 after AIDS diagnosis (49 cases between 12 and 4 months before AIDS diagnosis have been excluded). <sup>2</sup>Matched (1:1 for KS, 1:2 for NHL, 1:5 for other cancers) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South). <sup>3</sup>Cancers with no follow-up (date of death=date of diagnosis) have been excluded.

Rispetto alle tabelle precedenti, sono stati esclusi i seguenti casi e controlli:

a. I casi con tumore (49) diagnosticati tra 4 e 12 mesi prima dell'AIDS, per rendere minima la distorsione dovuta al "survival bias" descritto a pagina 4 (Figura 2) e, allo stesso tempo, limitare la perdita di casi dovuti alla concomitanza tra la diagnosi di AIDS e quella per tumore (Figura 3);

b. casi e controlli con follow-up pari a 0, cioè persone la cui data di diagnosi coincide con quella di morte.

La distribuzione per sesso di casi e controlli risulta identica (Tabella 5) in quanto nessun controllo di sesso diverso rispetto a quello dei casi è stato estratto. Differenze trascurabili emergono nella distribuzione per età, con una differenza massima nella mediana per i linfomi non-Hodgkin (36 nei casi e 39 nei controlli) e anno di diagnosi.

### **3.6 Follow-up**

La data comune di chiusura del follow-up è stata fissata al 31 dicembre 2008, data alla quale tutti i registri tumori inclusi nello studio garantivano un follow-up completo. Ciò consente di avere un sufficiente periodo di osservazione per calcolare la sopravvivenza a 3 anni dalla diagnosi con il metodo di “coorte”, cioè senza “censura”. Tale approccio (sopravvivenza di coorte) è necessario non potendo ipotizzare un “censoring non informativo” nel confronto tra i 2 gruppi di pazienti. Per ridurre le possibili perdite al follow-up dopo la diagnosi di AIDS, dovute al follow-up incompleto dello stato in vita effettuato dal registro nazionale AIDS, le informazioni sullo stato in vita dei PHA sono state recuperate attraverso un ulteriore record linkare con l’archivio nazionale di mortalità ISTAT.<sup>Serraino 2009</sup>

L’unico outcome considerato nel presente studio è la mortalità per tutte le cause, vista la difficoltà di classificare correttamente le cause di morte nei PHA.<sup>Serraino 2010, Zucchetto 2010</sup>

Per ciascun paziente con tumore, il rischio di morte è stato calcolato dalla data del tumore fino alla prima tra le seguenti date: data di morte, data di ultimo follow-up, 13/12/2008.

### **3.7 Analisi statistiche**

L’effetto della presenza di AIDS sul rischio di morte è stato stimato per mezzo di Death Hazard ratios (DH) attraverso un modello di rischio proporzionale di Cox<sup>Cox 1972</sup>. Le analisi sono state condotte utilizzando la regressione logistica condizionata (“individual matching”) implementata nella procedura PHREG di SAS. Un ulteriore aggiustamento per età e anno di calendario alla diagnosi (in anni) non modifica alcuna delle stime presentate.

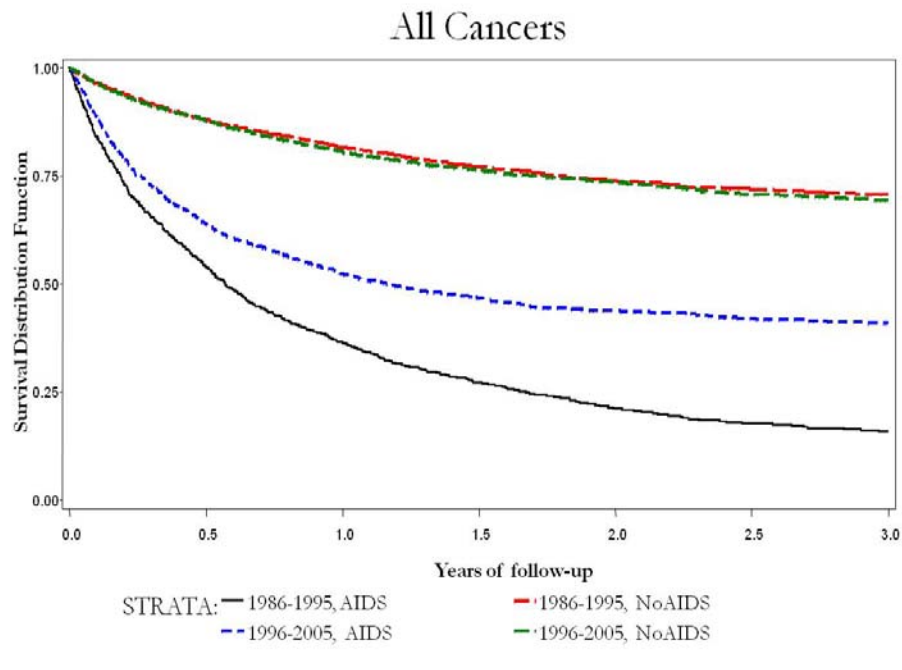
Sono state inoltre stimate le funzioni di distribuzione di sopravvivenza non parametriche<sup>Kaplan e Meier 1958</sup> attraverso la procedura LIFETEST di SAS in modo da confrontare (anche graficamente) le curve di sopravvivenza di due o più gruppi di pazienti.

## 4. RISULTATI DELLO STUDIO

### 4.1 Analisi generali per periodo

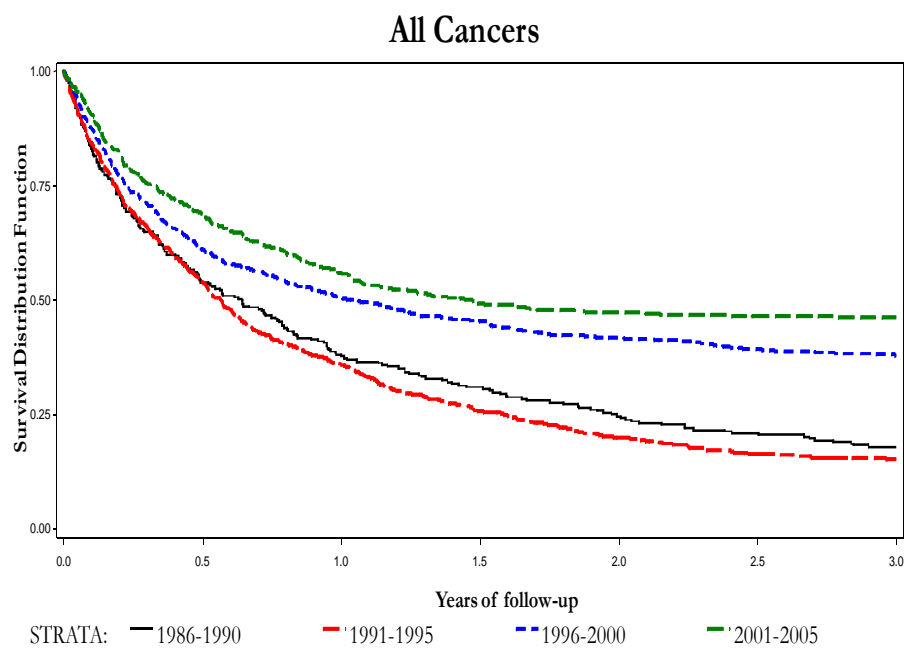
Per tutti i tumori nei PHA, la sopravvivenza a 3 anni è passata dal 16% del periodo 1986-1995 al 41% (Figura 4) nel periodo 1996-2005 (cART).

**Figura 4. Three-year Survival in persons with AIDS and matched controls by calendar period (1986-1995 vs 1996-2005)**



Le distribuzioni di sopravvivenza nei PHA per quinquennio di diagnosi mostrano un ulteriore significativo miglioramento (log-rank test,  $p < 0.01$ ) nel periodo 2001-2005 (46% dopo 3 anni), rispetto al 1996-2006 (38%) (Figura 5 e Tabella 6).

**Figura 5. Three-year Survival in persons with AIDS only  
by 5-years calendar period (1986-1990, 1991-1995, 1996-2000, 2001-2005)**



Nella Figura 4 e in tutte le figure successive, le curve relative ai pazienti con tumore senza AIDS (controlli) si riferiscono ai pazienti “appaiati” ai casi per sede tumorale, sesso, età, quinquennio di diagnosi e area di residenza. Di conseguenza, non rappresentano in alcun modo la sopravvivenza nella popolazione generale ma servono solo per valutare la sopravvivenza nei pazienti senza AIDS che presentano la stessa distribuzione delle principali covariate, rispetto ai tumori nei PHA.

Il confronto della sopravvivenza dopo il tumore tra PHA e pazienti senza AIDS è stato effettuato separatamente nei 2 periodi (1986-1995 e 1996-2005), visto l’impatto delle cART (Figura 4).

Nel periodo pre-cART (Tabella 6), anche i tumori a miglior prognosi nei pazienti senza AIDS (pelle, non melanomi e linfomi di Hodgkin), mostravano nei PHA sopravvivenze a 3 anni inferiori al 25% ed anche la sopravvivenza a 1 anno era molto limitata (36%) rispetto ai pazienti senza AIDS (82%) di pari età e sesso.

Il rischio di morire dopo un tumore (DH) per tutte le sedi neoplastiche era 5,5 volte più alto nei PHA e superiore a 5 per ciascuna singola sedi, a parte i carcinomi invasivi della cervice che mostravano un rischio pari a 3,2.

**Tabella 6. Three-year survival in persons with AIDS<sup>1</sup> and matched controls<sup>2</sup> and Death Hazards<sup>3</sup> in Italy, 1986-1995**

ICD10; Cancer type or site	No AIDS			AIDS			
	Cases	Deaths	Survival (SE) %	Cases	Deaths	Survival (SE) %	Death hazard (95% CI)
<b>ALL CANCERS</b>	<b>1899</b>	<b>556</b>	<b>70 (1)</b>	<b>988</b>	<b>831</b>	<b>16 (1)</b>	<b>5.5 (4.7-6.3)</b>
<b>AIDS-defining cancers</b>	<b>1156</b>	<b>436</b>	<b>72 (1)</b>	<b>918</b>	<b>767</b>	<b>16 (1)</b>	<b>5.1 (4.4-6.0)</b>
C46; Kaposi sarcoma	312	54	83 (2)	320	255	20 (2)	6.4 (4.4-9.1)
C82-85,88,96; NHL	1115	364	67 (1)	572	501	12 (1)	5.0 (4.2-6.0)
C53; Cervix uteri	129	18	86 (3)	26	11	58 (10)	3.2 (1.5-7.1)
<b>Non-AIDS-defining cancers</b>	<b>343</b>	<b>120</b>	<b>65 (3)</b>	<b>70</b>	<b>64</b>	<b>9 (3)</b>	<b>7.8 (5.1-11.9)</b>
C00-14, 30-32; Head&Neck	10	0	100 (-)	2	2	0 (-)	∞
C16; Stomach	15	9	40 (13)	3	3	0 (-)	∞
C18-20; Colorectum	29	11	62 (9)	6	6	0 (-)	∞
C21; Anus	15	3	80 (10)	3	3	0 (-)	∞
C22; Liver	9	7	22 (14)	2	2	0 (-)	∞
C33-34; Lung	39	28	28 (7)	8	8	0 (-)	∞
C44; Skin, non melanoma	20	2	90 (7)	4	3	25 (22)	∞
C70-72; Brain	19	10	47 (11)	4	4	0 (-)	∞
C81; Hodgkin lymphoma	94	8	91 (3)	19	15	21 (9)	39 (8-179)
C91-95; Leukaemias	15	9	40 (13)	3	3	0 (-)	∞
Other NADC	78	33	58 (6)	16	15	6 (6)	5.7 (2.4-11.9)

NOS=Not otherwise specified, CI=Confidence interval

<sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years after AIDS diagnosis.

<sup>2</sup>Matched (1:1 for KS, 1:2 for NHL, 1:5 for other cancers) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

<sup>3</sup>Estimates from the Cox proportional hazard model with stratification on matching variables as in footnote <sup>2</sup>.

**Tabella 7. Three-year survival in persons with AIDS<sup>1</sup> and matched controls<sup>2</sup> and Death Hazards<sup>3</sup> in Italy, 1996-2005**

ICD10; Cancer type or site	<i>No AIDS</i>			<i>AIDS</i>			
	Cases	Deaths	Survival (SE) %	Cases	Deaths	Survival (SE) %	Death hazard (95% CI)
<b>ALL CANCERS</b>	<b>2979</b>	<b>909</b>	<b>69 (1)</b>	<b>1212</b>	<b>714</b>	<b>41 (1)</b>	<b>3.0 (2.7-3.4)</b>
<b><i>AIDS-defining cancers</i></b>	<b>1904</b>	<b>483</b>	<b>75 (1)</b>	<b>995</b>	<b>579</b>	<b>42 (2)</b>	<b>3.3 (2.9-3.8)</b>
C46; Kaposi sarcoma	247	43	83 (2)	252	87	65 (3)	2.1 (1.4-3.1)
C82-85,88,96; NHL	1334	384	71 (1)	678	468	31 (2)	3.6 (3.1-4.3)
C53; Cervix uteri	323	56	83 (2)	65	24	63 (6)	2.4 (1.5-4.0)
<b><i>Non-AIDS-defining cancers</i></b>	<b>1075</b>	<b>426</b>	<b>60 (1)</b>	<b>217</b>	<b>135</b>	<b>38 (3)</b>	<b>2.6 (2.1-3.3)</b>
C00-14,C30-32; Head&Neck	35	16	54 (8)	7	6	14 (13)	3.4 (1.1-10.5)
C16; Stomach	40	16	60 (8)	8	6	25 (15)	1.8 (0.6-5.3)
C18-20; Colorectum	70	19	72 (5)	14	9	36 (13)	3.6 (1.5-8.7)
C21; Anus	49	13	73 (6)	10	7	30 (14)	4.0 (1.3-12.2)
C22; Liver	63	43	32 (6)	13	12	8 (7)	2.6 (1.2-5.9)
C33-C34; Trachea and lung	228	180	21 (3)	46	38	17 (6)	1.7 (1.1-2.5)
C44; Skin, non melanoma	140	8	94 (2)	28	7	75 (8)	5.0 (1.6-15.7)
C70-72; Brain	30	20	33 (9)	6	4	33 (19)	1.8 (0.5-6.1)
C81; Hodgkin lymphoma	177	22	88 (2)	36	19	47 (8)	8.6 (4.0-18.7)
C91-95; Leukaemias, all	30	10	67 (9)	6	6	0 (-)	∞
Other NADC	213	79	63 (3)	43	21	51 (8)	2.1 (1.2-3.5)

NOS=Not otherwise specified, CI=Confidence interval <sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years after AIDS diagnosis. <sup>2</sup>Matched (1:1 for KS, 1:2 for NHL, 1:5 for other cancers) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

<sup>3</sup>Estimates from the Cox proportional hazard model with stratification on matching variables as in footnote <sup>2</sup>

La Tabella 7 mostra le sopravvivenze dopo il 1996 (periodo cART) e confronta il rischio di morte nei pazienti con e senza AIDS. Per tutti i tumori, le sopravvivenze a 3 anni risultano essere del 69% nei pazienti senza AIDS e del 41% nei PHA, per un rischio di morte 3 (intervallo di confidenza al 95%: 2,7-3,4) volte superiore in quest'ultimo gruppo. Per alcuni tumori, tuttavia, è stato possibile registrare sopravvivenze superiori o vicine al 50% dopo 3 anni dalla diagnosi: tumori della pelle (75%), sarcomi di Kaposi (65%), carcinomi invasivi della cervice (63%) e linfomi di Hodgkin (47%). Il rischio di morire dopo un tumore rimane particolarmente elevato nei PHA rispetto ai pazienti senza AIDS per le sedi a buona prognosi:

linfomi di Hodgkin (DH=8,6), tumori della pelle non-melanomi (DH=5,0) e dell'ano (DH=4,0). Al contrario, per i tumori a prognosi severa (polmone e cervello) il rischio di morte nei PHA è 2 volte inferiore rispetto a quello nei pazienti senza AIDS (Tabella 7).

La Figura 6 descrive i miglioramenti di sopravvivenza nei PHA per i tumori inclusi nella definizione clinica di AIDS. Nel periodo 1996-2005, la sopravvivenza a 3 anni dopo un sarcoma di Kaposi è arrivata al 65%, “solo” 18% in meno rispetto ai pazienti senza AIDS di pari età, sesso e residenza (DH=2,1; Tabella 7). La differenza tra PHA e senza AIDS rimane invece molto marcata per la sopravvivenza dei linfomi non-Hodgkin (anche dopo aver “matchato” i controlli per istotipo – Tabella 3), rispettivamente 31% e 71% pari a un DH=3,6. Il miglioramento nei 2 periodi è invece trascurabile (non significativo) per i carcinomi invasivi della cervice, con una sopravvivenza a 3 anni nei PHA intorno al 60% e un rischio rispetto ai pazienti senza AIDS intorno a 3.

La distribuzione di sopravvivenza per le sedi tumorali non definitive di AIDS con almeno 5 casi nel periodo cART sono mostrate nella Figura 7. Per alcune sedi (stomaco, polmone, pelle non-melanomi e cervello), la sopravvivenza a 3 anni dal tumore è diventata molto simile nei pazienti con e senza AIDS. Differenze molto grandi rimangono tra i PHA, rispetto ai pazienti senza AIDS, per la sopravvivenza dei tumori di testa e collo (rispettivamente 14% e 54%), del colon retto (36% e 72%), dell'ano (30% e 73%), del fegato (8% e 32%), dei linfomi di Hodgkin (47% e 88%) e delle leucemie (0% e 67%) (Tabella 7 e Figura 7).



Figura 6. Three-year Survival in persons with AIDS and matched controls by calendar period (1986-1995 vs 1996-2005) for AIDS-defining cancers

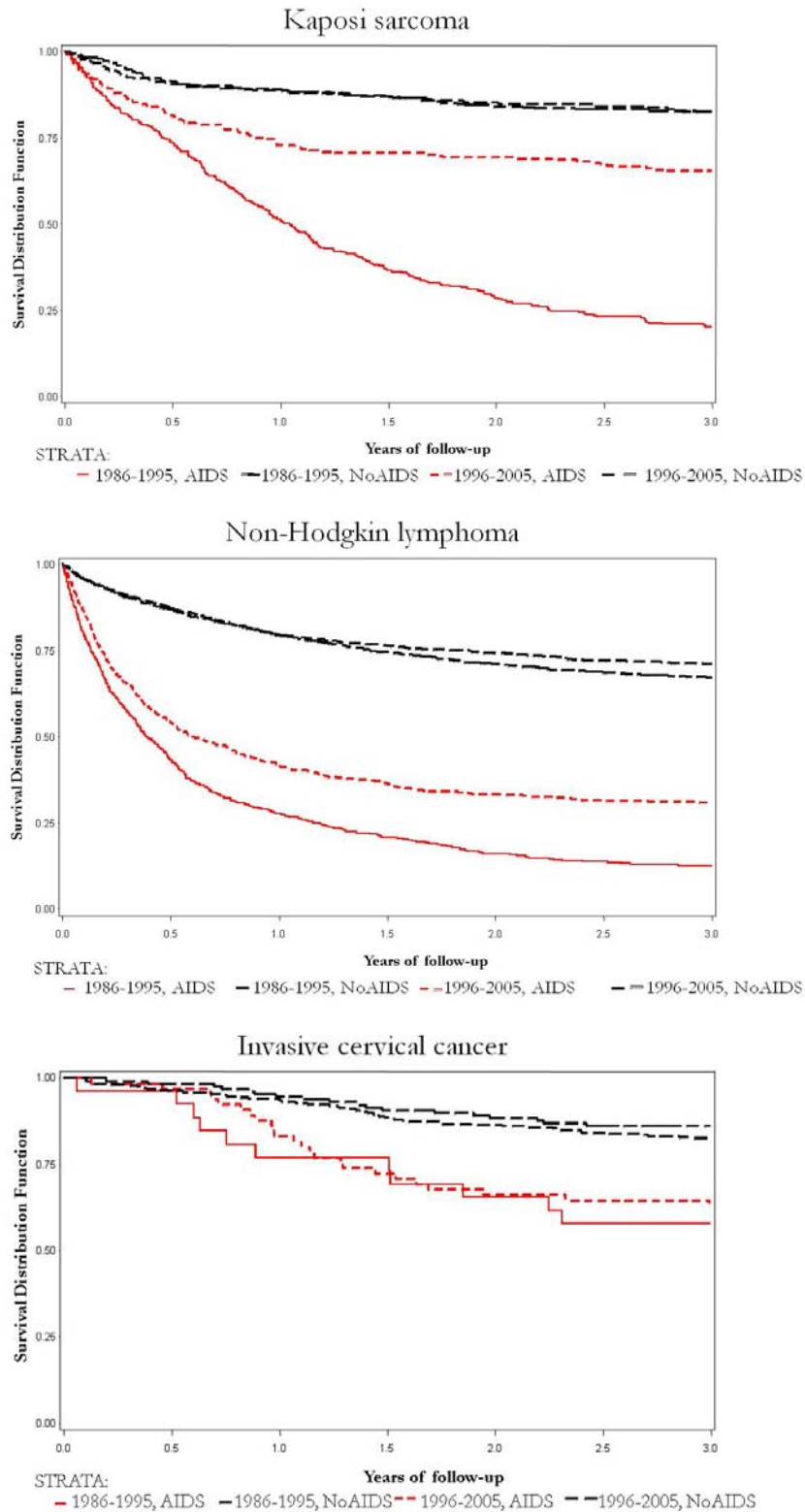


Figure 7. Three-year Survival in persons with AIDS and matched controls by calendar period (1986-1995 vs 1996-2005) for non-AIDS-defining cancers

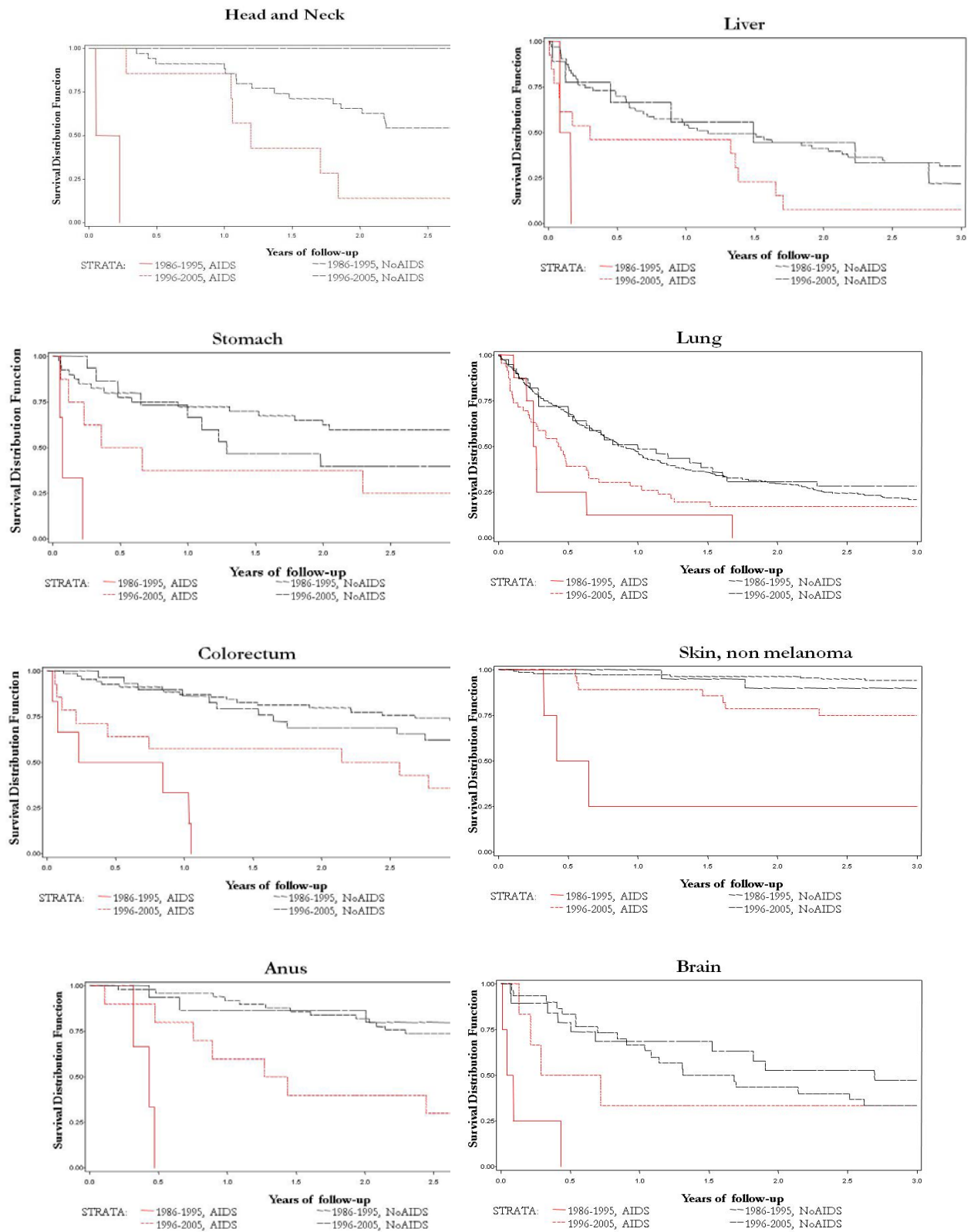
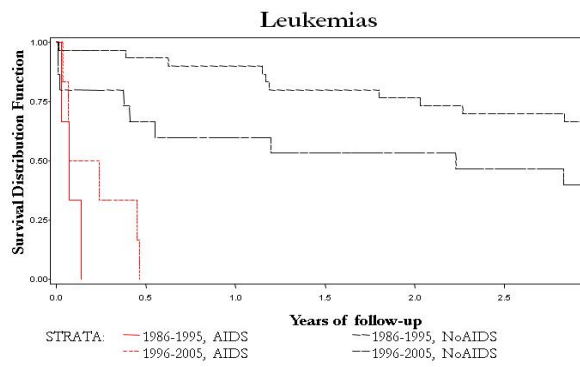
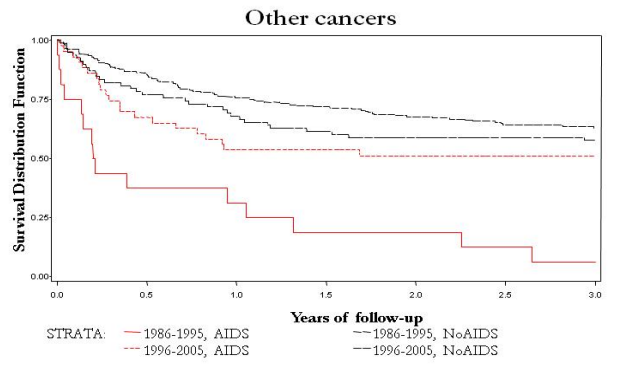
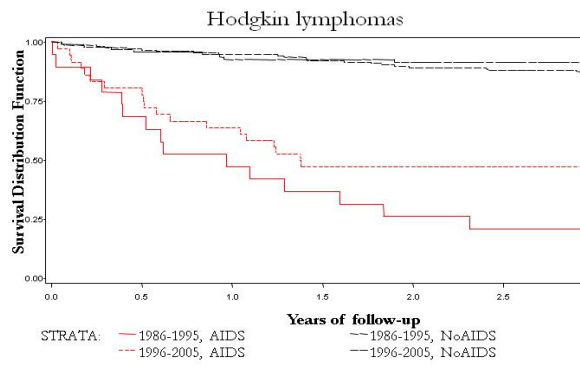


Figura 7 (continua).



## 4.2 Analisi stratificate

L'analisi stratificata, condotta per tutte le sedi tumorali nel periodo 1996-2005, non ha evidenziato differenze nel rischio di morte a seconda del sesso del paziente (3,1 nei maschi e 3,2 nelle femmine) (Tabella 8). Sono molto diversi, invece, i rischi di morte a seconda della categoria di trasmissione dei PHA, con un rischio di 2,4 per i PHA con tumore che hanno contratto l'infezione attraverso rapporti eterosessuali, 3,1 per i maschi omosessuali e di 3,8 nei PHA utilizzatori di droghe per via iniettiva. Quest'ultimo gruppo mostra anche una sopravvivenza a 3 anni (33%) significativamente inferiore a quella degli altri 2 gruppi (45 e 46%) (Figura 8).

**Tabella 8. Three-year Survival in persons with AIDS<sup>1</sup> and matched controls<sup>2</sup> and Death Hazards<sup>3</sup> by sex and transmission category in Italy, 1996-2005**

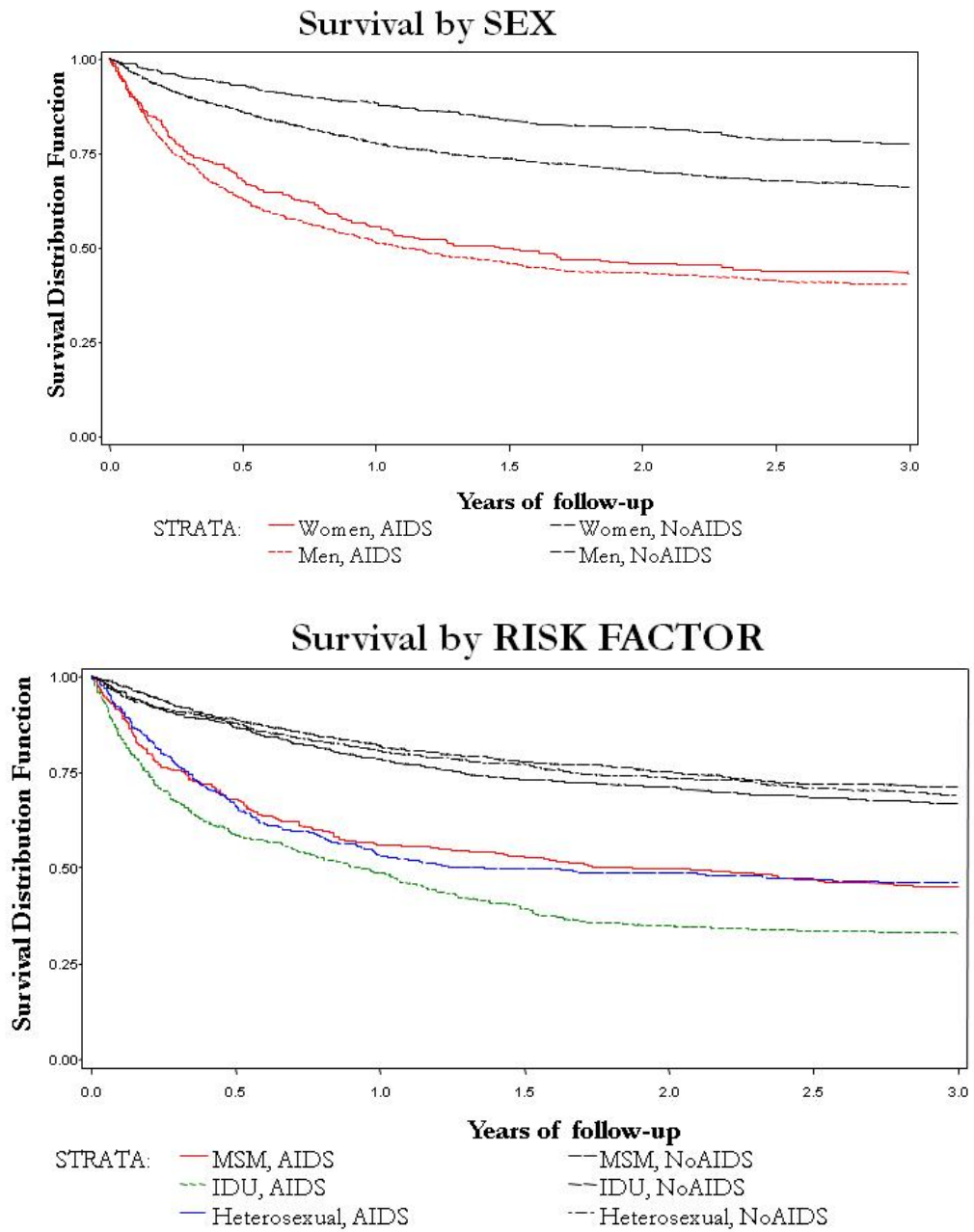
ICD10; Cancer type or site	<i>No AIDS</i>			<i>AIDS</i>			
	Cases	Deaths	Survival (SE) %	Cases	Deaths	Survival (SE) %	Death hazard (95% CI)
<b>ALL CANCERS</b>	<b>2979</b>	<b>909</b>	<b>69 (1)</b>	<b>1212</b>	<b>714</b>	<b>41 (1)</b>	<b>3.0 (2.7-3.4)</b>
Men	2156	725	78 (1)	946	563	43 (3)	3.1 (2.4-3.5)
Women	823	184	66 (1)	266	151	40 (2)	3.2 (2.5-4.1)
Men having sex with men <sup>4</sup>	707	233	67 (2)	351	192	45 (3)	3.1 (2.4-3.9)
Injecting drug users <sup>4</sup>	1252	361	71 (1)	447	300	33 (2)	3.8 (3.1-4.6)
Heterosexual <sup>4</sup>	1020	315	69 (1)	414	222	46 (2)	2.4 (2.0-3.0)

NOS=Not otherwise specified, CI=Confidence interval <sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years after AIDS diagnosis. <sup>2</sup>Matched (1:1 for KS, 1:2 for NHL, 1:5 for other cancers) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

<sup>3</sup>Estimates from the Cox proportional hazard model with stratification on matching variables as in footnote <sup>2</sup>

<sup>4</sup> No "risk factor" is available, control group represents matched patients according to stratification on matching variables <sup>2</sup>.

Figura 8. Three-year Survival in persons with AIDS and matched controls for all AIDS-defining cancers by sex and risk factors in Italy, 1996-2005



### 4.3 Analisi di sensitività

La Tabella 9 mostra i risultati principali di una serie di analisi di sensitività, al fine di valutare l'impatto sulla sopravvivenza dell'utilizzo di disegni di studio o gruppi di confronto alternativi.

Il rischio di morte non si modifica in modo rilevante se, come gruppo di controllo (anziché i pazienti accoppiati per sede, età, sesso e periodo di diagnosi), viene utilizzata l'intera coorte da cui sono stati estratti i controlli stessi (non rappresentativa dell'intera popolazione con tumori in Italia).

L'inclusione nel modello di regressione di aggiustamenti per le variabili di matching porta consistentemente ad una sovrastima del rischio di morte, con HR=4,2 per tutti i tumori non definitori rispetto a 2,6 ottenuto attraverso le stime "individual mathed" (Tabella 9). Risultati analoghi emergono per i linfomi non-Hodgkin (HR=5,4 e 3,6), meno evidenti per i sarcomi di Kaposi (2,4 e 2,1).

**Tabella 9. Death Hazards in persons with AIDS compared to cancer cases without AIDS according to different matching criteria and adjustment models in Italy, 1996-2005**

ICD10; Cancer type or site	<i>No AIDS</i>		<i>AIDS</i>			
	Cases	Deaths	Cases	Deaths	Model	Death hazard (95% CI)
<b>All non-AIDS-defining cancers</b>						
	<b>1075</b>	<b>426</b>	<b>217</b>	<b>135</b>	<b>Individual matching<sup>2</sup></b>	<b>2.6 (2.1-3.2)</b>
All cohort	353,923	124,612	217	135	No matching, unadjusted	2.4 (2.1-2.9)
Only >3 months after AIDS	353,923	124,612	151	91	Adj sex, age, year,area	4.2 (3.6-4.9)
					„	4.1 (3.4-5.6)
<b>C82-85,88,96; NHL</b>						
	<b>1334</b>	<b>384</b>	<b>678</b>	<b>468</b>	<b>Individual matching<sup>2</sup></b>	<b>3.6 (3.1-4.3)</b>
All cohort	23814	7086	678	468	No matching, unadjusted	3.7 (3.4-4.1)
Only >3 months after AIDS	23814	7086	140	112	Adj sex, age, year,area	5.4 (4.9-6.0)
Excluded cases at RAIDS only			486	384	“	7.6 (6.3-9.3)
					Individual matching <sup>2</sup>	3.6 (3.0-4.4)
<b>C46; Kaposi sarcoma</b>						
	<b>247</b>	<b>43</b>	<b>252</b>	<b>87</b>	<b>Individual matching<sup>2</sup></b>	<b>2.1 (1.4-3.1)</b>
All cohort	964	141	609	180	No matching, unadjusted	2.2 (1.8-2.8)
Only >3 months after AIDS	964	141	87	42	Adj sex, age, year,area	2.4 (1.8-3.2)
Excluded cases at RAIDS only			161	60	„	3.8 (2.5-5.8)
					Individual matching <sup>2</sup>	2.0 (1.3-3.2)

NOS=Not otherwise specified, CI=Confidence interval, NHL=non-Hodgkin lymphoma

<sup>1</sup>Cancers notified at CRs in people with AIDS, age 16 – 74 years after AIDS diagnosis.

<sup>2</sup>Matched (1:1 for KS, 1:2 for NHL, 1:5 for other cancers) for site (icd10), sex, age (quinquennia), period of diagnosis (1986-1990, 1991-1995, 1996-2000, 2001-2005); macroarea (North, Center, South).

<sup>3</sup>Estimates from the Cox proportional hazard model with stratification as in footnote <sup>2</sup>

L'esclusione di tutti i casi di tumore fino a 3 mesi dopo l'AIDS porta ad un'ulteriore sovrastima del rischio di morte, rispetto alle stime effettuate includendo i tumori da 3 mesi prima dell'AIDS (che abbiamo supposto ne anticipino la diagnosi), HR=4,1 per tutti i tumori, 7,6 per i linfomi non-Hodgkin (rispetto a 3,6) e 3,8 per i sarcomi di Kaposi (rispetto a 2,1). È interessante notare che, utilizzando questa finestra di osservazione che include solo i casi in una fase più avanzata dell'AIDS, la perdita di casi di tumore è stata di circa un quarto per i tumori non definitivi, metà per i sarcomi di Kaposi e quasi tre quarti per i linfomi non-Hodgkin.

In questo studio sono stati inclusi anche i tumori definitivi diagnosticati solamente al RAIDS, a differenza degli studi sull'eccesso di rischio in cui, per confrontare i tumori in AIDS con quelli della popolazione generale, il primo gruppo doveva necessariamente essere incluso nel secondo<sup>Dal Maso 2009</sup>. L'esclusione di questi tumori (18% dei sarcomi e 31% dei linfomi) non mostra nessun impatto sulle stime "individual matched" (Tabella 9). Un impatto considerevole si osserva escludendo i carcinomi invasivi della cervice diagnosticati solo al RAIDS, con stime di morte che diventano 5,2 rispetto a quelle calcolate (2,6) con l'intero gruppo di pazienti (dati non mostrati in Tabella 9). Tali discrepanze suggeriscono una mortalità differenziale, con la possibilità che almeno una parte dei casi diagnosticati solo al RAIDS (non rintracciati in alcun registro tumori) siano stati invece lesioni preinvasive della cervice a miglior prognosi.

## 5. DISCUSSIONE

Il rischio di morire dopo un tumore era 3 volte più alto nei PHA rispetto ai pazienti senza AIDS in Italia nel periodo 1996-2005, nonostante il recupero immunitario indotto dalle cART. La differenza nella sopravvivenza a 3 anni nei due gruppi di pazienti rimane di oltre il 20% (46% tra i PHA e 69% per i pazienti senza AIDS). Lo studio ha, tuttavia, mostrato che la sopravvivenza 3 anni dopo un tumore nei PHA è molto aumentata nell'era cART e incrementi si vedono anche nell'ultimo quinquennio osservato 2001-2005 (46%), rispetto al precedente (38%).

Questi risultati dovrebbero incoraggiare i clinici a migliorare la prognosi nei PHA attraverso procedure che portino ad una diagnosi più precoce e al trattamento di questi pazienti. La diagnosi di tumori nei PHA, pone l'oncologo di fronte a scelte difficili per bilanciare rischi e benefici nell'uso di farmaci (chemioterapici) potenzialmente immunosoppressivi.<sup>Mounier 2006</sup> I pazienti con tumore a migliore prognosi mostrano comunque una differenza di sopravvivenza rispetto ai pazienti senza AIDS e supportano le decisioni di utilizzare i trattamenti ottimali per i tumori quando le cART sono efficaci nel mantenere un

ragionevole livello di immunità. Approcci terapeutici aggressivi nei PHA possono essere considerati anche per i tumori a prognosi severa.<sup>Spina 2005; Sparano 2010</sup>

Nonostante il recupero immunitario che si verifica nei PHA trattati con cART, l'eccesso di rischio di sviluppare un tumore nei PHA è circa 3 volte superiore (2 per i tumori non definitivi) rispetto alla popolazione generale<sup>Dal Maso 2009; Franceschi 2010</sup> e la più bassa sopravvivenza di questi pazienti (mostrata in questo studio) suggeriscono la necessità di intensificare i programmi di prevenzione neoplastica nelle persone infette da HIV. In particolare, risultano urgenti programmi per la riduzione dell'abitudine al fumo ed il consumo di alcol, oltre ai trattamenti dell'infezione da HCV e delle lesioni preneoplastiche legate all'HPV.

Nel presente percorso di studio sono stati affrontati i problemi metodologici che si incontrano quando si cerca di stimare e confrontare incidenze e sopravvivenze tenendo conto della natura della sindrome da immunodeficienza acquisita. La definizione di caso AIDS ed il suo significato clinico è cambiato nel corso dei decenni e dei Paesi<sup>Kaldor 2008</sup> rendendo difficilmente confrontabili nel tempo e nello spazio le sopravvivenze (così come le incidenze) dei tumori dopo l'AIDS. Lo studio dell'epidemiologia dei tumori nella fase avanzata del decorso clinico dei pazienti con HIV (AIDS) attraverso l'utilizzo di disegni di studio tradizionali (es.: coorte di pazienti diagnosticati con AIDS, includendo solo gli eventi verificatisi dopo alcuni mesi dall'arruolamento) non consente, quindi, di stimare in modo accurato il fenomeno. Questo principalmente a causa del fatto che le stime di incidenza dei tumori e di rischio di morte dopo la neoplasia mostrano valori più elevati nel periodo post-AIDS.<sup>Dal Maso 2010b; Simard 2010</sup>

La stima della sopravvivenza calcolata con il metodo di coorte utilizzato nel presente lavoro come negli studi di sopravvivenza di popolazione<sup>AIRT 2007</sup> può sottostimare la "reale" attesa di sopravvivenza se, nei pazienti con diagnosi recente, sono subentrati fattori in grado di migliorare la sopravvivenza (diagnosi precoce, terapia). D'altra parte, il metodo evita assunzioni arbitrarie, non essendo realistico ipotizzare un "censoring non informativo" nel confronto tra i 2 gruppi di pazienti.

Il disegno "population based", che include tutti i casi di tumore nelle aree coperte da registri tumori, consente di effettuare stime rappresentative della sopravvivenza per l'intera popolazione con AIDS in Italia.

L'unico studio analogo, che confrontava i PHA a tutti i pazienti senza AIDS<sup>Biggar 2005</sup>, condotto negli Stati Uniti includeva casi fino al 2000, considerava i tumori in persone senza



AIDS tutti quelli che si verificano prima della diagnosi (in quel Paese basata sulla conta dei CD4) della sindrome e calcolava i DH a 2 anni dalla diagnosi. Lo studio mostrava un rischio di morte complessivamente più basso rispetto allo studio italiano presentato nella presente relazione. Il DH era, per i tumori definitivi, di 1,9 per i NHL nel periodo 1996-2000 (3,6 nel presente studio per il 1996-2005) e 1,8 per i tumori invasivi della cervice uterina (2,4 in Italia). Per gli altri tumori, il DH per il tumore del polmone era di 2,5 (1,7 in Italia), non significativamente aumentato per i tumori di colonretto e ano (1,6 e 0,9; 3,6 e 4,0 in Italia) e pari a 2,6 per i linfomi di Hodgkin (8,6 in Italia). I risultati dei due studi sono comunque difficilmente confrontabili a causa delle notevoli differenze metodologiche: utilizzo come gruppo di confronto dell'intera coorte e analisi aggiustate (non matchate), la diversa durata del follow-up (2 anni) e il diverso significato clinico della diagnosi di AIDS (causato dal criterio immunologico negli Stati Uniti, anziché clinico in Europa).

Il disegno di studio adottato (con gruppo di controllo "individual matched") ha consentito di stimare in modo non distorto il rischio di morte nei PHA. Tali stime sono state spesso effettuate includendo serie cliniche di pazienti<sup>Puoti 2004; Spagnolo 2011</sup> sia come casi che come gruppo di controllo, spesso arruolate in centri di eccellenza (selezionati), di conseguenza non rappresentative della sopravvivenza nei pazienti con o senza AIDS.

I risultati del presente Dottorato sono stati oggetti di alcune pubblicazioni, cui il dottorando ha collaborato, che hanno indagato l'intero spettro della storia naturale dei tumori nei PHA. In questi pazienti è stata stimata l'incidenza<sup>Polesel 2010</sup>, l'eccesso di rischio di tumore<sup>Dal Maso 2009, Franceschi 2010</sup>, la mortalità per tutte le cause neoplastiche<sup>Serraino 2010, Zucchetto 2010</sup> e la sopravvivenza (descritta nella presente relazione), sia in termini assoluti che rispetto ai pazienti senza AIDS, contribuendo ad una più solida comprensione del fenomeno.

## 6. APPENDICE

Descrizione dei programmi e delle macro SAS utilizzate per l'estrazione dei controlli, sviluppate dalla Dr.ssa Claudia Braga in collaborazione con il dottorando.

### REQUISITI DEL PROGRAMMA DI MATCHING (**MAIN\_MATCH.SAS**)

#### INFORMAZIONI NECESSARIE IN INPUT:

- a) Percorso e nome del dataset SAS sul quale si dovrà lavorare
- b) Nome della variabile caso/controllo (requisiti: numerica, 1 = caso, 2 = controllo)
- c) Nome della variabile identificativa dei soggetti (requisito: numerica, univoca)
- d) Nome della variabile tipo follow-up per la quale viene richiesto che il valore relativo al controllo matchato sia  $\geq$  a quello del caso corrispondente (requisiti: numerica). NB: andranno sistemati prima i casi con minore follow-up associandoli a un controlli con il minimo follow-up ( $\geq$ ).
- e) Numero dei controlli che si vorrebbero estrarre per ogni caso
- f) Numero delle variabili di stratificazione (requisiti: massimo 6 variabili)
- g) Per ogni variabile di stratificazione (requisiti: numeriche, tipo trend, valori ammessi fra 0 e 99: nel caso spetterà all'utente che lancia il programma provvedere a crearne una ri - codifica):
  - i. Nome della versione tipo trend della variabile sulla quale lavorare
  - ii. Nome della variabile contenente il valore non ricodificato (eventualmente uguale al precedente)
  - iii. Flag (0/1) che indichi se sono ammessi matching al di fuori della cella esatta per quella dimensione (verranno comunque effettuate ricerche spostandosi solo lungo una dimensione per volta: prima lungo la prima variabile dell'elenco per la quale si ammettano matchaggi al di fuori della cella esatta, poi lungo la seconda variabile di questo tipo, etc.)
- h) Path e nome dei dataset da creare come output (matchati e non matchati)

#### VINCOLI DEL PROGRAMMA E NOTE

1. Il programma scandirà prima tutti i casi cercando di matchare i controlli richiesti nella cella esattamente corrispondente, effettuerà poi in un secondo giro gli eventuali matchaggi nelle celle adiacenti quella esatta
2. Sarà ammesso uscire dalla cella esatta al massimo di una distanza 1, dal momento che si ritiene che eventuali matching mancanti debbano essere preferibilmente effettuati a mano
3. Il problema dei missing sulle variabili interessate dal programma dovrà essere trattato a monte dall'utente che lo lancia. Il programma segnalerà all'utente la presenza di missing e prima di cominciare l'elaborazione eliminerà tutti i record che li contengono
4. E' stata predisposta la macro **%MANMATCH** che, sulla base di un file Ascii (struttura record: "id caso; id controllo"), effettua i matchaggi manuali (con trasferimento nel file di output dei controlli e valorizzazione delle variabili che tengono traccia dell'esito del matching)

## RISULTATO DELLA ELABORAZIONE

1. Creazione di due dataset SAS: uno contenente tutti i casi e i controlli matchati a qualche caso, e uno contenente i controlli non matchati (nel caso si dovesse presentare la necessità di ripetere l'estrazione per qualche caso, sarà necessario rilanciare il programma su di un dataset appositamente ripredispuesto). In particolare, il primo di questi dataset dovrà contenere, oltre a tutte le variabili del dataset di lancio, le variabili seguenti:

a. **\_idMatch**: chiave di matching (per i casi: id del caso stesso, per i controlli: = id del caso cui sono matchati)

b. **\_FIMtcCa**: Codifica dell'esito del matching per i casi (per i controlli: missing, per i casi: = 0 se caso completamente matchato con controlli tutti reperiti nella cella esatta, = 1 se caso completamente matchato ma anche con controlli presi al di fuori della cella esatta, =2 se caso matchato completamente ma utilizzando anche dei matching selezionati manualmente, = 3 se caso non completamente o non matchato)

c. **\_FIMtcCo**:Codifica dell'esito del matching per i controlli (per i casi: missing, per i controlli: = 0 se controllo matchato nella cella esatta, = 1 se controllo matchato al di fuori della cella esatta, = 2 per i controlli matchati a mano (nella fase successiva manuale)

2. Calcolo di delle seguenti numerosità:

a. Numero casi matchati completamente nella cella esatta

b. Numero casi matchati andando a cercare anche al di fuori della cella esatta

c. Numero casi matchati utilizzando anche dei matching manuali (diverso da zero solo dopo avere utilizzato la macro **%manMatch**)

d. Numero casi non (completamente) matchati

3. Stampa dei seguenti gruppi di record (per i quali verranno stampate le seguenti informazioni: la variabile id, le variabili di matching (sia il valore effettivo che la categoria), la variabile di follow-up)

a. I casi matchati completamente con la lista dei corrispondenti controlli ordinati per: matchati nella cella esattamente corrispondente, matchati nelle celle adiacenti a quella esatta, matchati manualmente.

b. I casi non (completamente) matchati con la lista dei corrispondenti controlli (se ce ne sono)

4. Un file ASCII con, per ogni caso non (completamente) matchato, la lista dei possibili controlli fra i quali eventualmente scegliere dei soggetti da associate al caso. Tali soggetti sono selezionati con il criterio seguente: controlli con valori delle variabili di stratificazione che ammettano di uscire dalla cella esatta a distanza al massimo 1 da quella del caso (in questo caso si ammette di uscire anche in più dimensioni contemporaneamente) e valore della variabile di follow-up maggiore o uguale al 50% della lunghezza del follow-up del caso.

```

/*-----MAIN_MATCH.SAS-----*/
*   Main program for random matching cases to controls           *
*   *
*   Variables used by the macro (to be initialized in the main program) *
* LIBINP              = path of the input dataset                *
* DSINP              = name of the SAS dataset with cases and controls to be *
*                   matched                                     *
* LIBOUT             = path of the output dataset                *
* DSOUTMT           = name of the output SAS dataset to be created with *
*                   matching results                             *
* DSOUTNM           = name of the output SAS dataset to be created with *
*                   controls not matched to any case           *
* FLNXTCO           = name (with path) of the file were to save the list of *
*                   controls (for each case not (completely) matched) to *
*                   be analyzed by the operator to perform subsequently *
*                   manual matching                             *
*   *
* VARCACO           = name in the input dataset of the variable which stores *
*                   the case/control status (it should be: numeric with *
*                   1 to identify cases, 2 to identify controls. *
*                   No missing values allowed, otherwise the corresponding *
*                   record will be rejected)                    *
* VARID            = name in the input dataset of the variable which stores *
*                   the identification subjects number (it should be *
*                   numeric. *
*                   No missing values allowed, otherwise the corresponding *
*                   record will be rejected)                    *
* VARFUP           = name in the input dataset of the variable which stores *
*                   the time of follow-up (it should be numeric. *
*                   No missing values allowed, otherwise the corresponding *
*                   record will be rejected)                    *
*   *
* NMATCH           = number of controls to match to each case *
*   *
* NSTRVAR          = number of the stratification variables (at most 6 *
*                   variables) *
*   *
* VSTR&I           = (for each stratification variable) name of the *
*                   trend-like version (integer values between 0-99) of *
*                   the i-th stratification variable. *
*                   No missing values allowed, otherwise the corresponding *
*                   record will be rejected)                    *
* VSTR&I.OR       = (for each stratification variable) name of the *
*                   original version (before re-coding) of the i-th *
*                   stratification variable (otherwise equal to &&VSTR&I) *
* EXCEL&I         = (for each stratification variable) 0/1 (respectively) *
*                   if it is not/ it is possible *
*                   to match NEXT to the exact cell for the i-th *
*                   stratification variable (matchings next to the exact *
*                   cell (maximum distance: 1 cell) will be performed for *
*                   only one stratification variable at a time, *
*                   following the order with which variables are listed *
*                   in the macro variables declaration)        *
*   *
* CHECKS           = YES/NO respectively enables/disables additional *
*                   prints for check purposes                    *
*   *
*   *
*   *

```

```

* Notes:
* - Matchings will be performed in two stages:
*   during the first one, cases will be matched to controls found in the
*   exactly correspondent stratification cell and with the most similar
*   time of follow-up (the smallest value greater or equal);
*   during the second stage cases still not completely matched will be
*   matched to controls found in the next cells moving (one at a time)
*   along each stratification variable for which this operation is allowed
*   (&exCel = 1), and with the most similar time of follow-up).
* - When searching in the next stratification cells, the maximum distance
*   allowed will be 1, in consideration of the need of personal inspection
*   by the operator for more distant matchings.
* - Missings on variables needed by the matching procedure should be
*   treated before macro usage: observations with missings values on these
*   variables will be showed and discarded before matching
*
* Macro output:
* - The SAS dataset &libOut.&dsOutMt containing all records from cases and
*   the records from only controls matched to some case. This dataset will
*   contain all the variables of the input dataset and the three following
*   additional ones:
*   ° _idMatch = id of the case for cases, id of the matched case for
*               controls
*   ° _FlMtcCa = . for controls, for cases: 0 if case completely matched
*               with controls only from the exactly correspondent cell,
*               1 if case completely matched, but using also controls from
*               next to the exact cell, 2 if case completely matched, but
*               using also controls manually choosen, 3 if case not or
*               not completely matched
*   ° _FlMtcCo = . for cases, for controls: 0 if control matched in the
*               exactly correspondent cell, 1 if control matched
*               next to the exact cell, 2 will be used to code controls
*               manually matched by the operator
* - The SAS dataset &libOut.dsOutNM containing all the controls not matched
*   to any case.
* - A print of the following informations:
*   ° Number of cases completely matched in the exactly correspondent cell
*   ° Number of cases completely matched using also subjects from next to
*   exact cell
*   ° Number of cases completely matched using also manual matchings
*   ° Number of cases not or not completely matched
* - A print of the following groups of records:
*   ° cases completely matched with the corresponding controls
*   ° cases not (completely) matched with the corresponding controls (if
*   present)
* - The file &flNxtCo with the list of controls (for each case not
*   (completely) matched) to be analyzed by the operator to perform manual
*   matchings. Criteria for choosing such controls are the following:
*   ° a maximum distance of 1 from values of the stratification variables
*   for which is allowed to search also in the nearest cells (in this
*   case it will be allowed to move more than a dimation at a time)
*   ° same values of the stratification variables which do not allow
*   moving from the exact cell
*   ° a value of the follow-up variable greater or equal to the 50% of the
*   case's follow-up value
*****/

```

```

libname perm          'D:\DATAIN';
libname permout      'D:\DATAOUT';

proc contents data=perm.sede99; run;
proc sort data=perm.sede99; by etacan; run;
proc freq data=perm.sede99;
    Tables caco lfu icd10n sexgr sesanaN agegr etacan yeargr yeard macroarea macrogr;
run;

/*****FACCIO GIRARE IL PROGRAMMA PER L'ESTRAZIONE GENERALE 1:5*****/
/***** VANNO FATTI GIRARE I PRG MAIN_MATCH2011** PER LE SEDI PARTICOLARI :
        KS (1:1), NHL -sottotipi (1:2), ICC, pelle e HL (sottotipi 1:5)
*****/

%let libInp = d:\DATAIN;
%let dsInp  = sede99;

%let libOut = D:\DATAOUT;
%let dsOutMt = match_5_99;
%let dsOutNM = noMat_5_99;
%let flNxtCo = D:\DATAOUT\99_CasiNonMtc_e_PossibContr.txt;

%let varCaco = caco;
%let varId   = idnum;
%let varFup  = lfu;

%let nMatch = 5;

%let nStrVar = 5;

%let vStr1 = icd10n;
%let vStr1Or = icd10n;
%let exCel1 = 0;

%let vStr2 = sexgr;
%let vStr2Or = sesanaN;
%let exCel2 = 0;

%let vStr3 = agegr;
%let vStr3Or = etacan ;
%let exCel3 = 0;

%let vStr4 = yeargr;
%let vStr4Or = yeard;
%let exCel4 = 0;

%let vStr5 = macrogr;
%let vStr5Or = macroarea;
%let exCel5 = 1;

%let checks = SI;
%let dirCheck = D:\DATAOUT;

/*****
 * Options and macro invocation
 *****/
filename libmacro 'D:\PRG\';
options mautesource sasautos=libmacro /*mprint*/;

%Match;

```

## 7. BIBLIOGRAFIA

- AIRT Working Group. Italian cancer figures-report 2006: 1. Incidence, mortality and estimates. *Epidemiol Prev* 2006;30(1 Suppl 2):8–28. Available at: [www.registri-tumori.it](http://www.registri-tumori.it).
- AIRT Working Group. I tumori in Italia -- Rapporto 2011. La sopravvivenza dei pazienti oncologici in Italia. Cancer survival in Italy: Available at. <http://www.registri-tumori.it/cms/Rapp2011>.
- AIRTUM Working group. Italian cancer figures - Report 2010 Cancer prevalence in Italy: Persons living with cancer, long-term survivors, and cured patients. *Epidemiol Prev*, 2010, 34(5-6) Suppl.2: 1-188.
- Ajdacic-Gross V, Zellweger U, Wang J, Fleerackers Y, Somaini B. How complete is AIDS surveillance in Europe? An eagle eye comparison with mortality data. *J Epidemiol Commun Health* 2001;55:52–56.
- Ancelle-Park R. Expanded European AIDS case definition. *Lancet* 1993;341:441.
- Biggar RJ, Engels EA, Ly S, et al., “Survival after cancer diagnosis in persons with AIDS”, *J. Acquir. Immune. Defic. Syndr.* (2005), 39: pp. 293–299.
- Bräu N, Fox RK, Xiao P, et al., “Presentation and outcome of hepatocellular carcinoma in HIV-infected patients: a U.S.-Canadian multicenter study”, *J. Hepatol.* (2007), 47: pp. 527-537.
- Breslow NE, Day NE. *Statistical methods in cancer research. The design and analysis of cohort studies.* IARC scientific publication No. 82, vol. II. Lyon: International Agency for Research of Cancer, 1987.
- Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the Whole Cohort in the Analysis of Case-Cohort Data, *Am J Epidemiol* (2009), 169:1398–1405
- Centro Operativo AIDS. Update of notification of AIDS cases and new HIV infections in Italy. December 2007. *Notiziario dell’Istituto Superiore di Sanità.* 2008;21:3–23.
- Centro Operativo AIDS. Aggiornamento delle nuove diagnosi di infezione da HIV al 31 dicembre 2009 e dei casi di AIDS in Italia al 31 dicembre 2010. *Notiziario dell’Istituto Superiore di Sanità.* 2011;24(5):1–27.
- Chao C, Xu L, Abrams D, et al., “Survival of non-Hodgkin lymphoma patients with and without HIV infection in the era of combined antiretroviral therapy”, *AIDS.* (2010), 24: pp. 1765-1770.
- Clifford GM, Polesel J, Rickenbach M, et al., “Cancer risk in the Swiss HIV Cohort Study: associations with immunodeficiency, smoking, and highly active antiretroviral therapy”, *J. Natl. Cancer. Inst.* (2005), 97: pp. 425–432.
- Conti S, Farchi G, Galletti A, Masocco M, Napoli PA, Pezzotti P, et al. La notifica della mortalità per AIDS in Italia (1992): qualità della certificazione e sottonotifica. *Giornale Italiano dell’AIDS* 1997;8:12–16.
- Cox DR. Regression models and life-time tables. *J Royal Stat Soc B.* (1972);34:187-220.
- Curado MP, Edwards B, Shin HR, Ferlay J, Heanue M, Boyle P, et al (eds). *Cancer incidence in five continents.* IARC Sci Publ N. 160, vol. IX. Lyon: International Agency for Research of Cancer, 2008.
- Dal Maso L, APPENDIX 2. User’s Guide to SALI : Software for Record Linkage. In Ferretti S, Giacomini A (eds) *Cancer Registration Handbook* (2010a). Available at: <http://www.registri-tumori.it/cms/?q=node/1078>
- Dal Maso L, AIDS-related cancers. In: *Treatment Strategies-Oncology.* Holcroft R (Eds). The Cambridge Research Centre Ltd. London (2010b): pp 25-33. Available at: <http://treatment-strategies.co.uk/publications/oncology/>
- Dal Maso L, Braga C, Franceschi S. Methodology used for “software for automated linkage in Italy” (SALI). *J Biomed Inform* 2001a;34:387–95.
- Dal Maso L, Polesel J, Ascoli V, et al. for the Cancer and AIDS Registry Linkage (CARL) Study: Classic Kaposi’s sarcoma in Italy, 1985-98. *Br J Cancer*, 2005, 92(1): 188-193.

- Dal Maso L, Polesel J, Serraino D, et al., “Pattern of cancer in persons with AIDS in Italy in the HAART era”, *Br. J. Cancer.* (2009), 100:840-847.
- Dal Maso L, Rezza G, Zambon P, et al., “Non-Hodgkin lymphoma among young adults with and without AIDS in Italy”, *Int. J. Cancer.* (2001b), 93: pp.430-435.
- Engels EA, Brock MV, Chen J, et al, “Elevated incidence of lung cancer in HIV-infected individuals”, *J. Clin. Oncol.* (2006), 24: pp. 1383-1388.
- Franceschi S, Dal Maso L, Arniani S, Lo Re A, Barchielli A, Milandri C, et al. Linkage of AIDS and cancer registries in Italy. *Int J Cancer* 1998;75:831-4.
- Franceschi S, Lise M, Clifford GM, et al., “Changing patterns of cancer incidence in the early- and late-HAART periods: the Swiss HIV Cohort Study”, *Br. J. Cancer.* (2010), 103: pp. 416-422.
- Frisch M, Biggar RJ, Engels EA, Goedert JJ, for the AIDS-Cancer Match Registry Study Group. Association of cancer with AIDS-related immunosuppression in adults. *JAMA* 2001;285:1736–1745.
- Grulich AE, Wan X, Law MG, Coates M, Kaldor JM. Risk of cancer in people with AIDS. *AIDS* 1999;13:839–43.
- Harrison KM, Song R, Zhang X. “Life expectancy after HIV diagnosis based on national HIV surveillance data from 25 states, United States”, *J. Acquir. Immune. Defic. Syndr.* (2010), 53: pp. 124-130.
- Hessol NA, Pipkin S, Schwarcz S, et al., “The impact of highly active antiretroviral therapy on non-AIDS-defining cancers among adults with AIDS”, *Am. J. Epidemiol.* (2007), 165: pp. 1143-1153.
- Hymes KB, Cheung T, Greene JB, Prose NS, Marcus A, Ballard H, et al. Kaposi’s sarcoma in homosexual men: a report of eight cases. *Lancet* 1981;2:598-600.
- IARC. IARC monographs on the evaluation of carcinogenic risks to humans. Human immunodeficiency viruses and human T-cell lymphotropic viruses, vol. 67. Lyon: IARC Press, 1996.
- Kaldor JM, Delpech V, Guy RJ. “AIDS case reporting: do we still need it?”, *Lancet.* (2008), 373: pp. 181-183.
- Kaplan EL, Meier P, “Nonparametric Estimation from Incomplete Observations,” *Journal of the American Statistical Association;* (1958), 53: pp. 457–481.
- Li Y, Law M, McDonald A, Correll P, Kaldor JM, Grulich AE. Estimation of risk of cancers before occurrence of acquired immunodeficiency syndrome in persons infected with human immunodeficiency virus. *Am J Epidemiol* 2002;155:153–8.
- Lima VD, Hogg RS, Harrigan PR, Moore D, Yip B, Wood E, et al. Continued improvement in survival among HIV-infected individuals with newer forms of highly active antiretroviral therapy. *AIDS* 2007;21:685-92.
- Lodi S, Guiguet M, Costagliola D, et al. “Kaposi sarcoma incidence and survival among HIV-infected homosexual men after HIV seroconversion”, *J. Natl. Cancer. Inst.* (2010), 102: pp.784-792.
- Melbye MI, Coté TR, Kessler L, Gail MI, Biggar RJ and The AIDS/Cancer Working Group. High incidence of anal cancer among AIDS patients. *Lancet* 1994;343:636-639.
- Mounier N, Spina M, Gabarre J, et al., “AIDS-related non-Hodgkin lymphoma: final analysis of 485 patients treated with risk-adapted intensive chemotherapy”, *Blood* (2006), 107: pp. 3832–3840.
- Oehler-Janne C, Hugué F, Provncher S, et al. “HIV-specific differences in outcome of squamous cell carcinoma of the anal canal: a multicentric cohort study of HIV-positive patients receiving highly active antiretroviral therapy”, *J. Clin. Oncol.* (2008), 26:2550-2557.
- Polesel J, Franceschi S, Suligoi B, et al., “Cancer Incidence in People with AIDS”, *Int. J. Cancer.* (2010), 127: pp. 1437-1445.



- Puoti M, Bruno R, Soriano V, et al., "Hepatocellular carcinoma in HIV-infected patients: epidemiological features, clinical presentation and outcome", *AIDS* (2004), 18: pp. 2285–2293.
- Serraino D, De Paoli A, Zucchetto A, et al., "The impact of Kaposi sarcoma and non-Hodgkin lymphoma on mortality of people with AIDS in the highly active antiretroviral therapies era", *Cancer. Epidemiol.* (2010), 34: pp. 257-261.
- Serraino D, Zucchetto A, Suligo B, et al., "Survival after AIDS diagnosis in Italy, 1999–2006: a population-based study", *J. Acquir. Immune. Defic. Syndr.* (2009), 51: pp. 99–102.
- Simard EP, Pfeiffer RM, Engels EA. "Spectrum of cancer risk late after AIDS onset in the United States", *Arch. Intern. Med.* (2010), 170: pp. 1337-1345.
- Spagnuolo V, Galli L, Salpietro S, et al. "Ten-year survival among HIV-1 infected subjects with AIDS- or non-AIDS-defining malignancies." *Int J Cancer.* (2011). doi: 10.1002/ijc.26332.
- Sparano JA, Lee JY, Kaplan LD, et al., "Rituximab plus concurrent infusional EPOCH chemotherapy is highly effective in HIV-associated B-cell non-Hodgkin Lymphoma", *Blood.* (2010), 115: pp.3008-3016.
- Spina M, Jaeger U, Sparano JA, et al., "Rituximab plus infusional cyclophosphamide, doxorubicin, and etoposide (RCDE) in HIV-associated non-Hodgkin's lymphoma: pooled results from three phase II trials.", *Blood.* (2005), 105: pp. 1891–1897.
- World Health Organization. *International Classification of Diseases*, 10th revision. Geneva: WHO, 1992.
- Zucchetto A, Suligo B, De Paoli A, et al., "Excess mortality for non-AIDS-defining cancers among people with AIDS", *Clin. Infect. Dis.* (2010), 51(9): 1099-1101.