

I tumori in Italia - Rapporto 2009

Materiali e metodi

Italian cancer figures - Report 2009

Materials and methods

Materiali e Metodi

Sedi

Per la definizione delle sedi è stata utilizzata la Classificazione internazionale delle malattie 10^a revisione;¹ l'analisi è stata condotta per le principali sedi tumorali. Per le sedi 'tumore del polmone' e 'tumore della cervice uterina' è stato valutato anche il trend di incidenza delle due forme istologiche principali, i tumori di tipo squamoso e gli adenocarcinomi.

Incidenza

Con il termine incidenza si indicano le nuove diagnosi tumorali effettuate in una popolazione definita, in un certo arco temporale. Sono stati analizzati tutti i tumori maligni infiltranti. Per il tumore della vescica sono stati inclusi i casi maligni (C67), *in situ* (D09.0), benigni (D30.3) e incerti (D41.4).

Mortalità

Con il termine mortalità si indicano i decessi tumorali riscontrati in una popolazione definita, in un certo arco temporale. Sono stati utilizzati i dati di fonte Istat o ReNCaM (Registro nominativo delle cause di morte di Usl) per le stesse aree e gli stessi periodi di attività dei Registri tumori.

Tasso grezzo (di incidenza o mortalità)

Il tasso grezzo è una misura di frequenza che si costruisce riportando il numero di casi incidenti (o di decessi) per tumore (n), verificatisi in un certo intervallo di tempo (t) in una popolazione, e la popolazione nella quale sono stati diagnosticati (o deceduti) (N) per il tempo durante il quale sono avvenute le diagnosi (o i decessi) (t).

Tasso grezzo = numero di nuovi casi diagnosticati (o di decessi) in una popolazione in un definito intervallo temporale / Popolazione a rischio per il tempo in cui è stata osservata

Tasso grezzo = $[n/(N*t)]*100.000$

Di solito il tasso è riferito agli eventi occorsi in una popolazione in un anno. In questo modo la frazione esprime il numero di tumori (o di decessi) che si sono verificati per ogni soggetto della popolazione in un anno. Essendo i tumori un fenomeno relativamente raro, solo una piccola parte della popolazione si ammala (o muore) per tumore, e quindi la frazione *casi (o decessi) / popolazione a rischio di ammalarsi (o di morire) in un anno* è un numero decimale molto inferiore all'unità. Per rendere il numero più facilmente leggibile, i tassi di incidenza e di mortalità sono riferiti non a un singolo soggetto ma a 100.000 soggetti.

Tassi standardizzati per età

Una parte delle differenze tra i tassi osservati in aree diverse o in periodi diversi può essere dovuta alla diversa struttura per età delle relative popolazioni. I tumori sono infatti una

Materials and methods

Sites

We used the 10th revision of the International Classification of Diseases to define cancer sites;¹ we analysed the most relevant cancer sites. With regard to 'lung cancer' and 'cervix uteri cancer' sites, we evaluated the incidence of the two most frequent histological types: squamous cell and adenocarcinoma.

Incidence

Incidence is the number of newly diagnosed cancers that occur in a given population in a certain span of time. We analysed all malignant infiltrating cancers. With regard to bladder cancer we included infiltrating cancer (C67), in situ (D09.0), benign (D30.3) and borderline cancer (D41.4).

Mortality

Mortality is the number of cancer deaths that occur in a given population in a certain span of time.

We analysed data from local mortality registries (ReNCaM) for the same areas and periods of cancer registry activity.

Crude (incidence or mortality) rate

The crude rate is the number of incident cancer cases (or deaths) (n) that occur during a period of time (t) in a population, divided by the population where they have been diagnosed (or died) (N) and the time during which the diagnoses (or deaths) have occurred (t).

Crude rate = number of newly diagnosed cases (or deaths) in a population in a certain time span / population at risk for the time during which it is observed

Crude rate = $[n/(N*t)]*100,000$

Usually crude rates refer to cases and deaths during one year. Therefore, the fraction indicates the number of cases (or deaths) for each subject in the population during one year. Since cancers are rare, only a small proportion of subjects develop (or die from) a cancer, therefore the fraction "cases (or deaths) population at risk of being diagnosed with (or dying from) a cancer in a year" is a very small number, far smaller than 1. To make it easily readable, rates are multiplied by 100,000 and therefore indicate the number of cases or deaths every 100,000 subjects/year.

Age-standardized rates

Part of the differences we observed across registries may be due to the different age structure in the populations. Cancer is a disease particularly frequent among the elderly, therefore a population with a greater percentage of elderly people could have a greater number of cancer diagnoses. To make the comparison across populations possible, but also to compare the same population in different calendar years, we can use the method of age standardization.² With this method the incidence is referred to a population

patologia osservata prevalentemente nei soggetti in età avanzata, perciò in una popolazione con una quota maggiore di anziani si potrà riscontrare una maggior frequenza di tumori. Per rendere possibile il confronto si applica il metodo di *standardizzazione per età*,² tramite il quale l'incidenza viene riferita a una popolazione di struttura standard. In questo studio è stato utilizzato il metodo di *standardizzazione diretta*, impiegando la popolazione standard europea o la popolazione residente nelle aree AIRTUM del 1998, maschile e femminile, riportate in Tabella 1. Il tasso standardizzato indica quanti casi (o decessi) si osserverebbero se le popolazioni in questione avessero avuto la struttura per età della popolazione standard utilizzata.

Tasso standardizzato per età = Sommatoria dei prodotti dei tassi età specifici della popolazione in studio per i corrispondenti gruppi d'età della popolazione standard (P) / Sommatoria della popolazione standard (P)

$$\text{Tasso std} = [(\sum (n_i/N_i * t_i * P_i) / \sum P_i)] * 100,000$$

Selezione casistica

Nella presente monografia sono stati analizzati i *trend* recenti di incidenza e mortalità dei Registri tumori italiani.

I Registri italiani rilevano l'incidenza in una popolazione di quasi 19 milioni di abitanti, come riportato nel capitolo relativo alla descrizione della banca dati. Questa popolazione, maggiormente concentrata nel Nord Italia, non può essere considerata un campione *random* della popolazione nazionale. Allo stesso tempo, però, le informazioni raccolte dai registri costituiscono la principale fonte per la valutazione della patologia oncologica in termini di incidenza, mortalità e sopravvivenza. Per la stima degli andamenti recenti osservati nelle aree coperte da Registri tumori, è necessario tener conto del fatto che la registrazione nelle singole aree non è relativa allo stesso periodo: i Registri sono stati avviati in periodi differenti e presentano un diverso stato di aggiornamento (conseguente alla maggiore o minore tempestività nella produzione dei dati).

Una valutazione del trend complessivo, relativo cioè a un *pool* di Registri, richiede l'utilizzo di una popolazione omogenea all'interno dell'arco temporale considerato; è quindi opportuno includere nell'analisi i dati dei Registri tumori che coprono lo stesso periodo di registrazione, raggiungendo un compromesso tra rappresentatività geografica e omogeneità interna.

Il periodo di studio va dal 1998 al 2005. Sono stati utilizzati i dati dei registri tumori che, a maggio 2009, presentavano informazione completa per tutti gli otto anni considerati, o al più con due anni mancanti, ovvero venti registri di popolazione e due registri specializzati (il Registro del tumore della mammella femminile di Palermo e il Registro mesoteliomi della Liguria).

Al fine di condurre l'analisi su una matrice completa di dati (per registro e anno di calendario) è stata effettuata la stima

Age	Population AIRTUM 1998, Male	Population AIRTUM 1998, Female	European Population
0-4	46,175	40,885	80,000
5-9	47,267	42,247	70,000
10-14	47,489	42,283	70,000
15-19	52,886	47,301	70,000
20-24	68,410	62,044	70,000
25-29	82,029	74,438	70,000
30-34	87,588	79,017	70,000
35-39	79,963	73,065	70,000
40-44	70,835	65,751	70,000
45-49	67,861	64,355	70,000
50-54	67,556	64,762	70,000
55-59	64,039	63,380	60,000
60-64	59,171	60,920	50,000
65-69	53,004	58,559	40,000
70-74	44,697	55,781	30,000
75-79	31,935	46,951	20,000
80-84	15,572	26,805	10,000
85+	13,523	31,456	10,000
Total	1,000,000	1,000,000	1,000,000

Table 1. Standard populations used for computing age-standardized rates.

with a standard age structure. In this publication we used direct age-standardization with the European standard population or the male and female AIRTUM population of 1998, as reported in Table 1. The age-standardized rate indicates how many cases (or deaths) we would have had if the structure of the population we are studying had been that of the standard population.

Age-standardized rate = sum of the product of age-specific rates for the age classes of the standard population (P) / sum of the classes of the standard population (P)

$$\text{Standardized rate} = [(\sum (n_i/N_i * t_i * P_i) / \sum P_i)] * 100,000$$

Case selection

In this publication we analysed recent incidence and mortality trends of some Italian Cancer Registries (CRs).

Overall, the population covered by all CRs is about 19 million, one third of the whole Italian resident population, as described in the chapter on the AIRTUM database.

The registries have mainly been implemented in northern Italy, so we cannot consider this population a representative sample of the whole Italian population; however, information collected by CRs is the main source for evaluation of cancer incidence, mortality, and survival.

To estimate the recent overall trend we must take into account that registration does not cover the same calendar period in all the geographical areas involved: each CR we considered started to collect data in different years; moreover, each registry has different update status (due to different timeliness in data collection).

To obtain an efficient estimate of global trends we need to consider a homogeneous population within the analysed period, so we included registries with the same data availability. The struc-

dei casi incidenti relativa ai Registri e agli anni non disponibili. Per sei Registri tumori generali è stato necessario stimare il numero di casi incidenti negli anni 2004 e 2005, per tre Registri tumori generali la stima ha riguardato solo l'anno 2004, mentre per due Registri (uno di popolazione e uno specializzato) la procedura di stima è stata applicata all'anno 1998, come descritto in Tabella 2.

La numerosità della casistica analizzata tra il 1998 e il 2005 è pari a 818.017 casi (di cui il 12% stimati) e 342.444 decessi. Il 70% della popolazione risiede nel Nord Italia, il 17% nel Centro e il 13% al Sud.

Nelle schede relative a ogni singola sede tumorale sono riportati i risultati dell'analisi condotta sugli andamenti temporali nel periodo 1998-2005; la procedura adottata è la cosiddetta analisi *join-point*, che permette di quantificare le variazioni temporali e identificare eventuali punti di cambiamento del trend.

Sebbene l'obiettivo principale di questo lavoro sia la descrizione dei trend recenti, si è ritenuto opportuno valutare anche gli andamenti della patologia oncologica nel lungo periodo. Per far ciò sono stati selezionati i dati di otto Registri tumori relativi al periodo 1986-2005. Per uno di questi è stata effettuata la stima dei casi incidenti negli anni 1986 e 1987, per un altro dell'anno 1986, come descritto in Tabella 2. I risultati relativi ai trend di lungo periodo sono presentati in uno specifico capitolo (pp. 106).

ture of data forced us to compromise between the degree of geographic representativeness and homogeneity.

The period of study goes from 1998 to 2005. We used data from CRs with complete information (as of May 2009) or with at most two missing years. We included twenty general registries and two specialized registries (Breast CR of Palermo and Mesothelioma CR of Liguria). To analyse a complete data matrix (by registry and calendar year) we estimated the number of incident cases for the combinations of registries and years which were not available. We estimated incident cases in 2004 and 2005 for six CRs, only 2004 for three CRs, and for two CRs (a population registry and a specialized registry) we calculated the number of cases for the year 1998, as described in Table 2.

Between 1998 and 2005, 818,017 cases (12% estimated) and 342,444 deaths were analysed. About 70% of the analysed population lived in northern Italy, 17% in the centre and 13% in the south. We reported the results on recent trends in the cancer specific fact sheets: we used the join-point method, which makes it possible to quantify time variation and find changes in the trend.

Although the goal of this work is the analysis of recent incidence and mortality trends, we thought that an evaluation on the long-run trends would also be useful. Therefore we selected data from eight CRs for the period 1986-2005. We estimated the number of incident cases in 1986 and 1987 for one registry and in 1986 only for another registry, as reported in Table 2. A specific chapter describes the results of the long-run trend analysis.

	LONG-TERM TREND 1986-2005												RECENT TREND 1998-2005							
	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
Parma	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Firenze Prato	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ragusa	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Romagna	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	•
Torino	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	•
Genova	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	•	•
Veneto	•	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	•	•
Modena	•	•	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Ferrara													X	X	X	X	X	X	X	X
Umbria													X	X	X	X	X	X	X	X
F.V.G.													X	X	X	X	X	X	X	X
Reggio Emilia													X	X	X	X	X	X	X	X
Napoli													X	X	X	X	X	X	X	X
Sondrio													X	X	X	X	X	X	X	X
Siracusa													•	X	X	X	X	X	X	X
Trento													X	X	X	X	X	X	X	•
Latina													X	X	X	X	X	•	•	•
Sassari													X	X	X	X	X	•	•	•
Alto Adige													X	X	X	X	X	•	•	•
Salerno													X	X	X	X	X	•	•	•
Palermo Mammella													•	X	X	X	X	X	X	X
Liguria Mesoteliomi													X	X	X	X	X	X	X	X

• Estimate; X Observed.

Table 2. AIRTUM dataset. Available and estimated data for each year and registry for long-term (1986-2005) and recent trends (1998-2005).

Procedura di stima

Il metodo utilizzato, in accordo con quello applicato precedentemente nella valutazione dei trend AIRTUM 1986-1997,³ consiste nello stimare i casi incidenti nell'anno o due anni mancanti, per ogni registro, in base all'intera serie storica del registro stesso. Questo permette di usufruire delle informazioni relative ai differenti andamenti di ogni singola area.

La procedura di stima è stata eseguita per ogni Registro e per ognuna delle 36 sedi tumorali oggetto della monografia, distintamente per maschi e femmine. Per quanto concerne il tumore del polmone e della cervice uterina è stato calcolato anche il numero totale atteso per i due gruppi morfologici 'adenocarcinomi' e 'carcinomi squamosi'.

Il lavoro si è articolato in due fasi:

- è stata preliminarmente condotta un'analisi *join-point*⁴ sui tassi standardizzati (standard: popolazione europea) per ogni singolo registro, sede e sesso, relativamente a tutto il periodo disponibile nella banca dati. Le caratteristiche dell'analisi sono descritte nel paragrafo seguente;
- successivamente è stato stimato il numero di casi per l'anno (o due anni) 'mancanti' sulla base di un modello log-lineare adattato sul periodo temporale che costituisce l'ultimo (o il primo) segmento risultante dall'analisi *join-point*:

$$\log(\text{rate}_{ij}) = \beta_{\text{anno}} + \alpha_i \text{età}_i$$

dove l'età è espressa in classi quinquennali (0-5, 5-9, ... 85+). I casi sono stimati utilizzando un modello di regressione che assume la linearità dei trend (più precisamente del logaritmo dei tassi di incidenza e mortalità). Seppure l'assunzione di un trend lineare possa, per alcune sedi, apparire ingiustificata, è necessario considerare che ogni curva può essere approssimata localmente in modo soddisfacente con un segmento lineare, purché di lunghezza appropriata. Inoltre la restrizione dell'arco temporale (su cui adattare il modello) all'ultimo segmento dell'analisi *join-point* ha reso possibile produrre stime valide anche per tutte quelle sedi che presentavano variazioni dell'andamento nel periodo.

Analisi *join-point*

La stima del cambiamento percentuale medio annuo dei tassi di incidenza e mortalità è espressa dall'APC (*annual percent change*), un indicatore che fornisce una notazione sintetica del segno e dell'intensità della variazione temporale dei tassi.

Il modello impiegato si basa sull'individuazione dei segmenti lineari che meglio si adattano ai tassi osservati, ovvero minimizzano la somma dei quadrati delle differenze tra i dati stimati e osservati.

Il trend è scomposto in segmenti, il cui numero massimo è limitato dal numero di *join-point* (punti di giunzione) arbitrariamente fissato prima dell'analisi. Per k *join-point* il trend può essere rappresentato al massimo da $k+1$ segmenti con diversa pendenza. L'anno in cui si stima verificarsi il *join-point* è l'anno che individua una variazione del trend.

Estimation procedure

The method we adopted is in agreement with the one that was used in the evaluation of the 1986-1997 trends³. For each registry we estimated the number of incident cases for the missing one/two years, using all the available time series for the CR. Therefore, for each area, we were able to use the information on different time trends.

We estimated the number of cases for each registry and for the 36 cancer sites considered, separately for men and women. As regards to lung cancer and cervix uteri cancer we calculated the number of cases for the two most frequent histological types: squamous cell and adenocarcinoma.

The two main phases of this work were as follows:

- *fitting of a join-point model⁴ on age standardized rates (standard: European population) for the whole period taken into account, for each registry, site and sex. Join-point analysis is explained in the next paragraph.*
- *estimation of the number of cases for the missing one or two years: we used a log-linear model, fitted on the last (or the first) segment resulting by the join-point analysis:*

$$\log(\text{rate}_{ij}) = \beta_{\text{year}} + \alpha_i \text{age}_i$$

where age is compiled in 5-year classes (0-4, 5-9, ... 85+).

We used a regression model which assumes the linearity of the trend (of the logarithm of the incidence and mortality rates). The assumption of linearity may not always be justified. However, each trend may be considered linear in a sufficiently short period of time. Moreover, fitting the model on a shorter period (the last or the first segment resulting by the join-point analysis) makes it possible to give valid estimates even for cancer sites with significant variations in trends.

Join-point analysis

The average annual percent change of incidence and mortality rates is expressed by an indicator, called APC (annual percent change), which explains synthetically the sign and the intensity of the variation in time.

This model is based on linear segments connected at join-points that represent the best fit of the observed data, that is, the segments that minimize the sum of the square of the differences between estimated and observed data.

The time trend is divided into segments: the number of segments depends on the number of join-points, fixed arbitrarily a-priori. For k join-points we can have at most $k+1$ segments with different slopes. The join-point year is the point in time when we estimate the trend variation.

For the 1998-2005 trend we found at most two segments with different slopes (that is, none or one join-point). For the 1986-2005 period, we found at most four segments with different slopes (that is, at most three join-points).

We fixed two further constraints defining the model:

- *the number of observations between two join-points must be at least four;*

Per quanto riguarda il periodo 1998-2005, il trend è stato scomposto al più in due segmenti con diversa pendenza (ovvero massimo un punto di *join-point*). Per quanto riguarda il periodo 1986-2005, il trend è stato scomposto al più in quattro segmenti con diversa pendenza (ovvero massimo tre punti di *join-point*).

Sono stati previsti ulteriori vincoli in sede di definizione del modello:

■ il numero anni di osservazione tra due *join-point* deve essere almeno quattro;

■ il numero di anni di osservazione antecedenti al primo *join-point* (o successivi all'ultimo) deve essere almeno pari a tre.

Il modello log-lineare *join-point* può essere rappresentato in un'unica equazione:

$$\ln(\text{Rate}) = \beta_0 + \beta_1 \text{anno}_i + \delta_1 (\text{anno}_i - \tau_1)^+ + \delta_2 (\text{anno}_i - \tau_2)^+ + \varepsilon_i^{(k)}$$

dove $\exp(\beta_0)$ è il valore stimato del tasso (di incidenza o mortalità) per il primo anno e $\varepsilon_i^{(k)}$ è un termine di errore che rappresenta la variabilità casuale.

Nel caso di un modello privo di *join-point* l'equazione si riduce all'equazione di una retta con intercetta β_0 e pendenza β_1 . I termini tipo $\delta_i(a)^+$ rappresentano, invece, la variazione della pendenza della retta per gli eventuali successivi segmenti e sono uguali a zero negli anni antecedenti quello di *join-point*.

Il modello migliore è individuato mediante un algoritmo che prevede il confronto tra modelli con un numero diverso di *join-point*. Il primo confronto è effettuato tra il modello privo di *join-point* e quello con un numero di *join-point* pari al numero massimo fissato. Il modello finale è quello con il minor numero di parametri, per il quale l'aggiunta di un ulteriore parametro (punto di *join-point*) non comporti miglioramenti significativi ($\alpha = 0.05$ aggiustata per test multipli).

Una volta selezionato il modello migliore la misura della variazione annua percentuale (APC) stimata per ogni segmento è pari a:

$$\text{APC} = 100 \times (e^\beta - 1)$$

La stima degli intervalli di confidenza è basata sulle seguenti formule:

$$\text{LCI} = 100 \times (e^{\beta - (t [n-p; 0.025] \times \text{E.S.})} - 1)$$

$$\text{UCI} = 100 \times (e^{\beta + (t [n-p; 0.025] \times \text{E.S.})} - 1)$$

dove E.S. è l'errore standard di β e t è il 95° percentile della distribuzione t di Student per $(n-p)$ gradi di libertà, con $p = 2k + 2$. Nelle singole schede sono riportati i risultati dell'applicazione della procedura sia sui tassi standardizzati relativi a tutte le età, sia a specifiche fasce di età (0-49 anni, 50-69 anni e 70+).

Le analisi sono state condotte utilizzando il software Joinpoint Regression Program, Version 3.3.1 - April 2008; Statistical Research and Applications Branch, National Cancer Institute (<http://srab.cancer.gov/joinpoint/>).

■ the number of observations before the first *join-point* (or following the last one) must be at least three.

The *join-point* log-linear model may be described by the following equation:

$$\ln(\text{Rate}) = \beta_0 + \beta_1 \text{year}_i + \delta_1 (\text{year}_i - \tau_1)^+ + \delta_2 (\text{year}_i - \tau_2)^+ + \varepsilon_i^{(k)}$$

where $\exp(\beta_0)$ is the estimated incidence or mortality rate for the first year and $\varepsilon_i^{(k)}$ represents the random variability in the measures. In a model without *join-point* the equation is a straight line with constant term β_0 and slope β_1 . $\delta_i(a)^+$ is the change in slope for the second, or third segment, if there are any *join-points*; it is equal to zero for the years before the *join-point*.

Join-point analysis finds the best model using a specific algorithm; the algorithm compares models with different number of *join-points*. The first comparison is between the model with no *join-point* and the model with maximum number of *join-points* (fixed a-priori).

The final model is the model with the least number of parameters, for which no significant improvement is gained by adding further parameters (*join-points*), $\alpha = 0.05$, adjusted for multiple tests.

When the model is defined, it is possible to measure the APC for each line of the trend, according to the following formula:

$$\text{APC} = 100 \times (e^\beta - 1)$$

Confidence interval formulas are:

$$\text{LCI} = 100 \times (e^{\beta - (t [n-p; 0.025] \times \text{S.E.})} - 1)$$

$$\text{UCI} = 100 \times (e^{\beta + (t [n-p; 0.025] \times \text{S.E.})} - 1)$$

where S.E. is the standard error of β and t is the quantile of a Student's t distribution with $(n-p)$ degrees of freedom, with $p = 2k + 2$. In each cancer site fact sheet we reported results of this method's application on both age-standardized rates and rates standardized on specific age groups (0-49 years, 50-69 years, and 70+). We used the Joinpoint Regression Program software, Version 3.3 - April 2008; Statistical Research and Applications Branch, National Cancer Institute (<http://srab.cancer.gov/joinpoint/>)

References - Bibliografia

1. Ministero della Sanità. Classificazione Statistica Internazionale delle malattie e dei problemi correlati, 10^a revisione. Ginevra, Organizzazione Mondiale della Sanità, 2000.
2. Bray F. Age-standardization. In: Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB, (eds). *Cancer Incidence in Five Continents*. Lyon, IARC Scientific Publications n. 155, Vol. VIII, 2002, 87-89.
3. Crocetti E et al. Gli andamenti della patologia oncologica in Italia: i dati dei Registri Tumori (1986-1997). *Epidemiol Prev* 2004; 28(2) suppl: 1-112.
4. Kim H J, Fay M P, Feuer E J, Midthune D N. Permutation tests for joinpoint regression with applications to cancer rates, *Statistics in Medicine*, 2000; 19: 335-351.