

revista cubana de
**SALUD
PUBLICA**

Revista Cubana de Salud Pública
Sociedad Cubana de Administración de Salud
ecimed@infomed.sld.cu
ISSN (Versión impresa): 0864-3466
CUBA

2004

Arduino Verdecchia / Roberta De Angelis / Silvia Franceschi
A METHOD FOR ASCERTAINING THE QUALITY OF CANCER REGISTRY DATA
Revista Cubana de Salud Pública, abril-junio, año/vol. 30, número 002
Sociedad Cubana de Administración de Salud
La Habana, Cuba

Red de Revistas Científicas de América Latina y el Caribe, España y Portugal

Universidad Autónoma del Estado de México



Instituto Superiore di Sanità Rome, Italy

A method for ascertaining the quality of cancer registry data

Arduino Verdecchia,¹ Roberta De Angelis,¹ Silvia Franceschi¹

Summary

Methods commonly used to evaluate completeness and quality of CR data involve simple and indirect indicators such as the proportion of DCO cases, the proportion of cases with microscopic confirmation, the incidence to mortality ratio, etc. Modelling cancer incidence, mortality and survival in a unified framework is proposed as an appropriate approach to check for consistency CR data with a process that turns out to be circular if, and only if, mortality, incidence and patients' survival data are high quality and complete. The MIAMOD method was developed to provide incidence, prevalence and mortality estimates and projections, using mortality and patients' survival information at national or regional levels. Examples of application of the MIAMOD method to national cancer registries in Europe are given to show the performance of the method in checking the CR data for internal consistency and completeness of registration. We are proposing the MIAMOD method and software as a tool for CR useful to check for completeness and quality of their data and to provide future cancer burden information for health planning and allocation of resources for their area.

Key words: Epidemiological methods, Cancer, Incidence, Mortality.

Introduction

Cancer registries (CR) represent epidemiological instruments which are aimed at providing population based cancer incidence and mortality. Cancer registries need well recognized requisites to be classified as high quality CR, and to be efficient tools for cancer control. Quality and completeness of data collected by cancer registries represent key issues that affect their reliability and potential use of their data. *When cancer patients' survival and prevalence are additional aims for a CR, further control is needed on completeness and quality of follow-up of patients for life status.*

Methods commonly used to evaluate the completeness and quality of CR data involve simple and indirect indicators such as the proportion of DCO cases, the proportion of cases with microscopic confirmation, the mortality to incidence ratio, etc. However, none one of these indicators must be taken individually for ascertaining the completeness and quality of collected data. For example, the mortality to incidence ratio is not expected to be constant in time as it strongly depends on the shape of incidence and mortality trend curves, according to their stable, increasing, decreasing or changing patterns. The

proportion of DCO cases can certainly inform us about the completeness of the collecting system but it does not give any information on the quality of diagnoses, dates, and the entity of all information collected.

Modeling cancer incidence, mortality and survival in a unified framework is, in principle, an appropriate approach to check for consistency of CR data. This kind of models establishes the theoretical relationships between the different quantities involved in the process to which the same cancer patients contribute. We can check whether incidence is consistent with mortality and patients' survival in a process that turns out to be circular if, and only if all the data, are of high quality and highly complete.

The Mortality and Incidence Analysis MODEL (MIAMOD) method^{1,2} was developed to provide incidence, prevalence and mortality estimates and projections, using mortality and patients' survival information at national or regional levels the MIAMOD method can be applied to data of a CR that has operated for several calendar year (at least 8-10 years) and provides comparative incidence estimates that are the ones expected on the basis of cancer mortality and patients' survival. From the match of the observed and expected incidence rates we derive quality issues for that the specific CR.

Aim of this paper is to propose the MIAMOD method and software as a useful tool for CRs who wish to perform their own check for completeness and quality of their data. CRs can also use the method to produce estimates and projections of cancer incidence, mortality and prevalence for their area.

Methods

The MIAMOD method was extensively presented in previous papers.^{1,2} We will present here its basic formulation and explain understandably how this method performs.

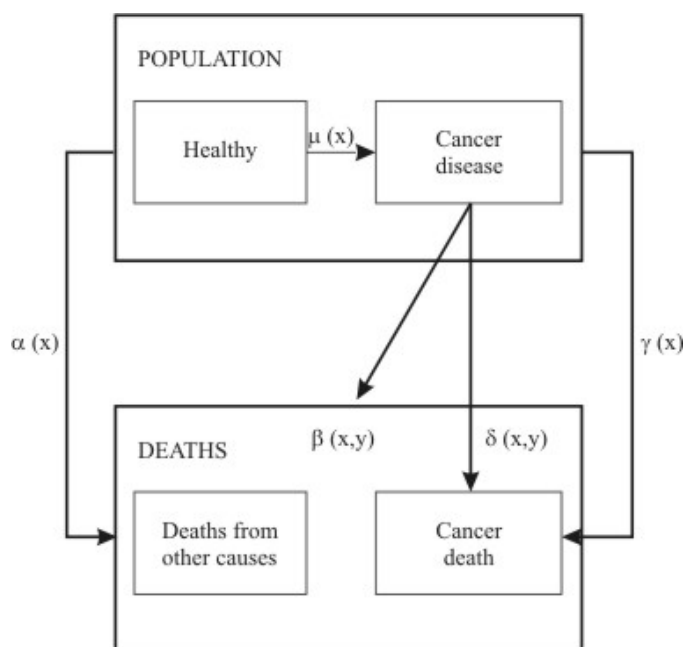


FIG. 1. The MIAMOD method: compartmental representation for a given birth cohort.

For a given birth cohort we can express cancer mortality $\gamma(x)$ at age x and prevalence $v(x)$ as

$$\gamma(x) = \int_0^x [1-v(\tau)] \mu(\tau) \delta(x, \tau) \exp\left\{-\int_{\tau}^x [\beta(u, \tau) - \alpha(u)] du\right\} d\tau \quad (1)$$

follows:

where $\mu(\tau)$, is cancer incidence, $\delta(x, \tau)$ is the cancer death hazard at age x for patients diagnosed

$$v(x) = \int_0^x [1-\tau(\tau)] \mu(\tau) \exp\left\{-\int_{\tau}^x [\beta(u, \tau) - \alpha(u)] du\right\} d\tau \quad (2)$$

with cancer at age τ , and β and α are the death hazard for cancer patients and death hazard for the general population, respectively. Equation (1) gives the expected cancer mortality at age x as the convolution of the probability of cancer diagnosis $\mu(\tau)$ at age $\tau < x$, for the proportion of healthy people in the population, $[1-v(\tau)]$, times the probability of death from cancer at age x , conditional to have survived the extra death hazard for cancer patients with respect to the general population.

The second equation gives similarly the cancer prevalence $n(x)$ at age x as the convolution the probability of cancer diagnosis $\mu(\tau)$ at age $\tau < x$, for the proportion of healthy people in the population, times the probability of surviving the extra death hazard for cancer patients with respect to the general population. A set of equation systems (1 and 2), one for each $N=A+Y$ birth cohorts involved in the data for A single year of age, and Y single year of diagnosis, allows functionally for a link function $G(\mu)$ between cancer incidence and mortality. Figure 2 show the data space for the MIAMOD method, including the back projection, estimation and projection areas. For each birth cohort the data matrix describes only a variable part of it and some back projection and forward projection is implied in the age period and cohort approach. In the projection area no data are available for the future birth cohort and a limit to 10-20 years of projection period should be defined in order to moderate the progressive bias of missing future birth cohorts.

Incidence is assumed as a polynomial function of age, period of diagnosis and birth cohort, throughout a logit link function Φ :

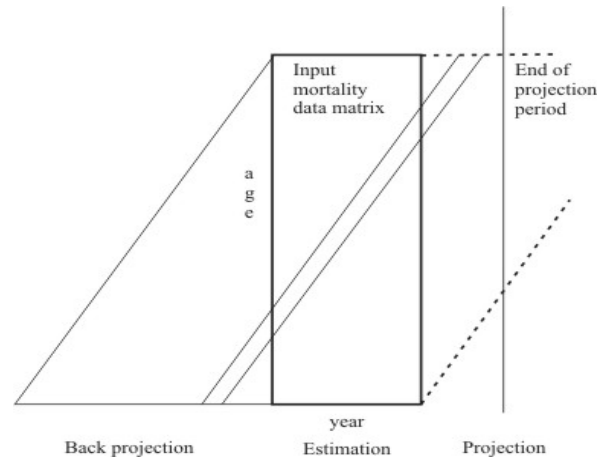


FIG. 2. The MIAMOD method data space.

$$\Phi(\mu(x, t, \Theta)) = const + \sum_{k=1}^{k_1} a_k x^k + \sum_{k=1}^{k_2} b_k t^k + \sum_{k=1}^{k_3} c_k (t - x)^k$$

where $\theta = (const, a_1, \dots, a_{k_1}, b_1, \dots, b_{k_2}, c_1, \dots, c_{k_3})$ is the vector of the parameters to be estimated by a maximum likelihood fit of cancer mortality data matrix by single year of age and calendar year of diagnosis. The function Φ works as a link function between incidence and mortality. Also the degree of polynomials have to be estimated. A set of restricted cubic splines³ can be used to model either age, period and cohort as an alternative to polynomials.

The MIAMOD method receives as an input age specific mortality data for a set of calendar years, for a specific cancer site of interest, age specific all causes mortality and population size for the same calendar years, and an estimate of patients' survival by age and calendar year. The MIAMOD method furnishes expected incidence, mortality and prevalence, with projections to a chosen projection period.

In a CR data application, whether or not the expected incidence matches the observed one, we can derive indications on the completeness and quality of CR data. If observed and expected data match we simply conclude that the CR data area high quality and complete. If they do not match we have to further study in detail all the information involved and to discriminate which of the information is likely to be problematic. Hence we must try to identify the ways to improve the data. The lack of consistency between incidence, mortality and survival can derive from problems of (in)completeness of cancer registration, misclassification of incidence or deaths cases. If a registry misses randomly a proportion of cases, the estimated patients' survival is not expected to be biased. In this case the lack of match of the observed and expected incidence expresses the proportion of missing incidence cases. If misclassification of the cancer site occurs frequently for collected incidence cases, the estimated patients' survival will reflect this biased disease definition of incidence cases that is expect to apply differently to mortality data. The lack of match between the observed and expected incidence in this case expresses the combined action of incidence misclassification and the corresponding patients' survival bias. Then, we expect that the observed and estimated cancer incidence will match perfectly if and only

if all the involved data are complete and high quality.

The method needs this patients' survival information over a long span of calendar years in order to cover potentially the entire data space (Figure 2). Survival data are usually available from CRs for a limited time period which depend on how long the CR has been operating. Some modeling of survival is then needed to expand the survival information both to long term and backward to former calendar years.

Cure models with covariates⁴⁻⁵ are used to model relative survival by age class and period of diagnosis to allow to expand survival as needed, according to some simple hypotheses i.e. constant, linearly increasing or decreasing, etc.

Example applications

The MIAMOD method was extensively applied to national level in Italy.⁶⁻¹⁰ to European countries within a concerted action of the EU commission, the EUROPREVAL Study, to the major cancer sites (Franceschi S, De Angelis R, Quinn M, Colonna M, Verdecchia A. Changing trend in lung cancer in Europa [in press] in and to US (Verdecchia A, Mariotto A, Micheli A, Ries LAG, Lynch CF, Yancik R. Estimating and projecting the prevalence of cancer: an Application to Iowa Cancer Registry data [in press]). The method was also applied to rather small areas, e.g. Iceland, local CR areas,¹¹ Italian regions,¹²⁻¹⁴ and rare cancer sites.

We will present in this section some example application from the EUROPREVAL experience which illustrate how the method performs and may be used to check CR data for their consistency.

Application to high quality CR

Figure 3 shows the estimated mortality and incidence in comparison to the observed mortality and incidence example applications to high quality cancer registries. The method fits observed mortality and reconstructs the expected incidence that is consistent with the levels and trends of mortality and patients' survival. The comparison of observed and estimated mortality serves as a goodness of fit evaluation while the comparison of observed and estimated incidence is simply a match that informs us about the consistency of all the involved data. Goodness of fit is generally high since the polynomial age, period and cohort model is rather flexible. The perfect matches that we individuate for prostate cancer incidence in Sweden, colorectal cancer incidence in Finland and Estonia assure us of the completeness and high quality of these CR data. For Iceland the huge variability of observed incidence and mortality data makes it difficult to ascertain clearly the levels and trends. Notwithstanding, the estimated incidence helps us to identify an increasing trend for colorectal cancer incidence that is rather consistent with the variable observed rates.

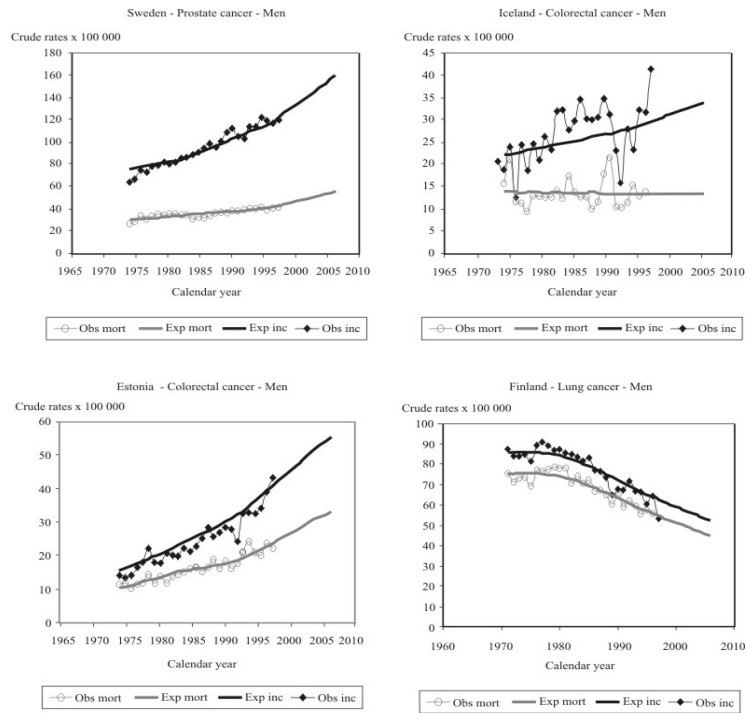


FIG. 3. Example applications to high quality national CR a) Prostate cancer in Sweden; b) Male colorectal cancer in Iceland; c) Male colorectal in Estonia; d) Male lung cancer in Finland.

Managing inconsistencies

When the comparison of observed and estimated incidence results in no satisfactory match, some further study is required to identify the reasons for these inconsistencies. We present some example situations with identified problems.

Figure 4 shows observed and estimated male lung cancer incidence and mortality in Sweden.

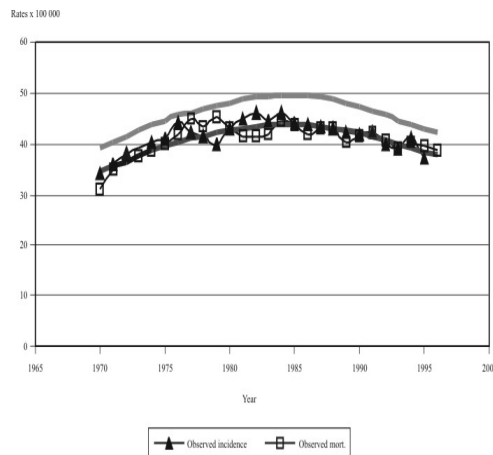


FIG. 4. Male lung cancer in Sweden. Comparison of observed and estimated incidence and mortality rates.

Looking at observed rates (black diamond marker for incidence and empty square marker for mortality) we clearly see that incidence and mortality overlap each other. This situation is consistent with a null patients' survival probability. Patients' survival from lung cancer in fact is poor (about 10 % of survivors at 5 years from the diagnosis) but not null. Estimated lung cancer incidence trends do not match at all with the observed data. The reason for this inconsistency can be easily found in the choice of the Swedish Cancer Registry, not to include the Death Certificate Only (DCO) cases in their incidence statistics. Sweden publishes yearly on a separate report a description of its DCO cases. Once the observed incidence is corrected with the DCO cases (see Table) in fact, the estimated (MIAMOD) and observed (EUROCIM-corrected) rates become very close to each other.

Figure 5 shows observed and estimated incidence and mortality trends for breast cancer in Scotland. Breast cancer incidence in Scotland presents a bump just following the year 1990 that is not reflected in the mortality trend and not reproduced by the estimated incidence. This is a very clear illustration of the effect, which the introduction of the population based breast cancer screening had on England and Scotland in 1990.¹⁵ After the bump, the incidence trend continued exactly as expected by the MIAMOD method. So in this case, the MIAMOD application served to quantify the artifact incidence effect of the screening we can evaluate by the area of the bump.

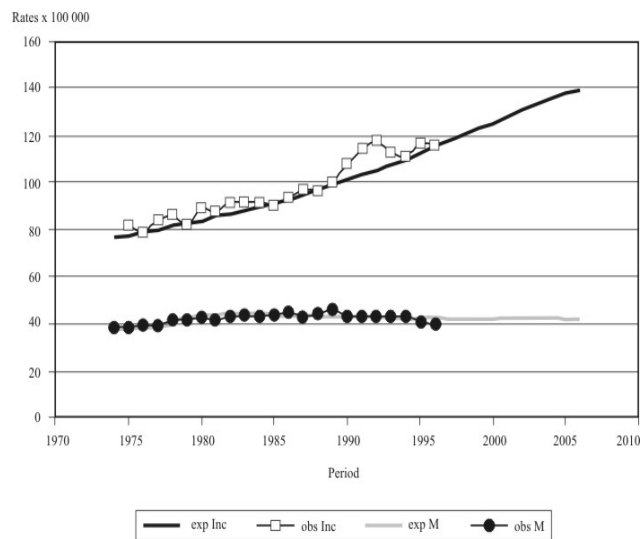


FIG. 5. Female breast cancer in Scotland. Comparison of observed and estimated incidence and mortality rates.

TABLE. Lung cancer in Sweden. Comparison of corrected observed incidence (EUROCIM-rates[§]) and MIAMOD estimated incidence (MIAMOD- rates[§])

a) MEN

	DCO cases			EUROCIM			MIAMOD		
	0-49	50-74	75+	ALL	cases	Rate [§]	cases*	Rates [§] *	Rates [§]
1996	5	47	119	171	1569	36	1740	40	42

1997	2	61	108	171	1605	37	1776	41	42
1998	1	67	137	205	1614	37	1819	42	41
a) WOMEN									
	DCO cases			EUROCIM			MIAMOD		
	0-49	50-74	75 +	ALL	cases	rates§	cases*	rates§*	rates§
1996	1	40	76	117	1026	24	1143	26	28
1997	2	39	88	129	983	23	1112	26	29
1998	0	36	66	102	1006	13	1108	26	30

§ Rates per 100,000 population.

* Corrected with the number of DCO cases.

Discussion

We proved that the MIAMOD method is a flexible and valid procedure to produce estimates and projections of incidence and prevalence, provided that valid mortality and survival information is available. The method can be also applied to rather small CR areas.

When used to check and validate CR data the method allows for a comprehensive evaluation of the consistency of incidence, mortality and patients' survival information that is integrative to other conventional methods currently used for data quality check. The MIAMOD application is not simply a check of the data as it heavily involves the analysis and use of the data. The application may result in a perfect match between the observed and estimated cancer incidence on the basis that all data are high quality and complete. For problems arising from an unsatisfactory match between observed and estimated cancer incidence we need to study in detail all possible reasons, i.e. patients selections, inconsistencies in the disease definition between incidence and mortality or other misclassifications, biases in the survival analysis and modeling, screenings, etc. Managing inconsistencies involves further studies on the quality and completeness of the CR data that certainly have the effect of improving the quality of the data.

Application of the MIAMOD method to CR areas can provide cancer prevalence estimates. Prevalence is an important indicator of the cancer burden in a population as it describes the size of the population with a previous diagnosis of cancer. This population constitutes a large part of the health demand in terms of costs for main treatment, palliative care, follow-up for recurrences, etc. Total prevalence includes all the patients who had a previous diagnosis of cancer, irrespectively to time since diagnosis, whether treated, cured or not. Partial prevalence estimates by time since diagnosis can be obtained from the MIAMOD application to identify strata of the prevalence that are more homogeneous in terms of same or similar care needs, i.e. for colorectal cancer, 1 year prevalence is to include patients under major treatment, whereas 5 year prevalence includes patients requiring major care needs and 5+ prevalence

includes patients potentially cured from the disease and therefore with a care demand is expected to be far smaller than average. When we derive prevalence numerically from CR data,¹⁶ invariably we refer always to the past. With the aim to provide useful information for public health, health planning, the allocation of financial and health care resources, we need to provide prevalence figures for coming years, and not the past. The MIAMOD method can allow a CR to provide administrators, practitioners, health structures and general public with information on the future cancer burden in their area.

The MIAMOD software, equipped with a very user friendly interface is available, free of charge, on request to Dr. Roberta De Angelis (e-mail: rodeange@iss.it). Using the method correctly, however requires some statistical and modelling knowledge, in addition to these basic features explained.

A first introductory course on modeling cancer incidence and mortality with MIAMOD will be organized in Italy, early December 2003. Other courses will follow on the basis of interest. Readers interested in further information on the MIAMOD software and courses may feel free to contact Dr. Roberta De Angelis, email: rodeange@iss.it.

RESUMEN

Los métodos que habitualmente se utilizan para evaluar el completamiento y la calidad de las informaciones del Registro del Cáncer comprenden indicadores sencillos e indirectos como son la proporción de casos con certificados de defunción (casos DCO), la proporción de casos con confirmación microscópica, la relación incidencia-mortalidad, etc. Se propone en este trabajo la modelación de la incidencia del cáncer, la mortalidad y la supervivencia en un marco unificado como método adecuado para chequear la consistencia de las informaciones del Registro del Cáncer con un proceso que se cierra solo cuando los datos sobre mortalidad, incidencia y supervivencia del paciente son completos y tienen una gran calidad. El método MIAMOD se desarrolló para brindar estimados y proyecciones referidos a la incidencia, prevalencia y mortalidad, utilizando la información sobre mortalidad y supervivencia de los pacientes a nivel nacional o regional. Se ofrecen ejemplos de la aplicación del método MIAMOD en los registros nacionales del cáncer en Europa para mostrar el comportamiento del método en la revisión de las informaciones del Registro del Cáncer en lo que respecta a la consistencia interna y el completamiento de la información. Proponemos el método y el software MIAMOD como una herramienta útil para chequear el completamiento y la calidad de las mencionadas informaciones y para brindar información futura sobre el peso del cáncer en los planes de salud y en la asignación de recursos para esta esfera.

Palabras clave: Metodos epidemiologicos, cAncer, incidencia, mortalidad.

Referencias bibliográficas

1. Verdecchia A, Capocaccia R, Egidi V, Golini A. A method for the estimation of chronic disease morbidity and trend from mortality data. *Stat Med* 1989;8: 201-16.
2. De Angelis G, De Angelis R, Frova L, Verdecchia A. MIAMOD: a computer program to

- estimate chronic disease morbidity using mortality and survival data. *Comput Methods Programs Biomed* 1994;44: 99-107.
3. Durrleman S, Simon R. 'Flexible Regression Models with Cubic Splines'. *Stat Medicine* 1989; 8:551-61.
 4. Verdecchia A, De Angelis R, Capocaccia R, Sant M, Micheli A, Gatta G, et al. The cure of colon cancer: results from the Eurocare Study. *Int J Cancer* 1998;77: 322-29.
 5. De Angelis R, Capocaccia R, Hakulinen T, Soderman B, Verdecchia A. Mixture models for cancer survival analysis: application to population-based data with covariates. *Stat Med* 1999; 18: 441-54.
 6. Capocaccia R, Verdecchia A, Micheli A, Sant M, Gatta G, Berrino F. Breast cancer incidence and prevalence estimated from survival and mortality. *Cancer Causes Control* 1990;1:23-30.
 7. Capocaccia R, Micheli A, Berrino F, Gatta G, Sant M, Ruzza MR, et al. Time trends of lung and larynx cancers in Italy. *Int J Cancer* 1994;57:1-8.
 8. Capocaccia R, De Angelis R, Frova L, Sant M, Buratti E, Gatta G, et al. Estimation and projections of stomach cancer trends in Italy. *Cancer Causes Control* 1995;6: 339-46.
 9. Capocaccia R, De Angelis R, Frova L, Gatta G, Sant M, Micheli A, et al. Estimation and projections of colorectal cancer trends in Italy. *Int J Epidem* 1997;26: 924-32.
 10. Verdecchia A, Mariotto A, Capocaccia R, Gatta G, Micheli A, Sant M, et al. Incidence and prevalence of all cancerous diseases in Italy: trends and implications. *Eur J Cancer* 2001;37: 1149-57.
 11. Micheli A. Cancer prevalence in Italy: the ITAPREVAL Study. *Tumori* 1999; 85:400-7.
 12. Micheli A, Verdecchia A, Capocaccia R, De Angelis G, Gatta G, Sant M, et al. Estimated incidence and prevalence of female breast cancer in Italian Regions. *Tumori* 1992;78:13-21.
 13. De Angelis R, Valente F, Frova L, Verdecchia A, Gatta G, Chessa E, et al. Trends of colorectal cancer incidence and prevalence in Italian Regions. *Tumori* 1998;84:1-8.
 14. De Angelis R, Valente F, Frova L, Capocaccia R, Micheli A, Chessa E, et al. Incidence, mortality and prevalence of stomach cancer in Italian Regions. *Tumori* 1996; 82:314-20.
 15. Effect of NHS Breast Cancer Screening Programme on Mortality from Breast Cancer in England and Wales, 1990-8: Comparison of Observed with Predicted Mortality. *BMJ* 2000;320(35): 665-69.
 16. Krogh V, Micheli A. A measure of cancer prevalence with a computerized program: an example of larynx cancer. *Tumori* 1996;82:1-4.

Recibido: 10 de septiembre de 2003. Aprobado: 11 de diciembre de 2003.

Arduino Verdecchia. Instituto Superiore di Sanità Lab. Epidemiología e Biostatistica. Reparto Indicatori sorveglianza sanitaria. Viale Regina Elena 299, 00161 Roma e-mail:verdeck@iss.it

¹ **Medical Doctor. Epidemiology y Bioestatistic.**

[Indice Anterior](#) [Siguiente](#)