ESTIMATED AND OBSERVED CANCER INCIDENCE IN ITALY: A VALIDATION STUDY

Riccardo Capocaccia¹, Carlotta Buzzoni², Enrico Grande¹, Riccardo Inghelmann¹, Francesco Bellù³, Tiziana Cassetti⁴, Margherita de Dottori⁵, Andrea Donato⁶, Vincenzo De Lisi⁷, Fabio Falcini⁸, Massimo Federico⁹, Stefano Ferretti¹⁰, Mario Fusco¹¹, Adriano Giacomini¹², Stefano Guzzinati¹³, Lucia Mangone¹⁴, Silvano Piffer¹⁵, Stefano Rosso¹⁶, Ornella Sechi¹⁷, Giovanna Tagliabue¹⁸, Rosario Tumino¹⁹, Marina Vercelli²⁰, and Susanna Vitarelli²¹

¹Reparto di Epidemiologia dei Tumori, Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità, Rome; ²Centro per lo Studio e la Prevenzione Oncologica, Regione Toscana, Florence; ³Registro Tumori dell'Alto Adige - Tumorregister Sudtirol, Ospedale di Bolzano, Bolzano; ⁴Registro Tumori Umbro di Popolazione, Università di Perugia, Perugia; ⁵Registro Tumori del Friuli-Venezia Giulia, Agenzia regionale della sanità, Udine; ⁶Registro Tumori della Provincia di Salerno, Salerno; ⁷Registro Tumori della Provincia di Parma, UO Oncologia, Agenzia ospedaliera, Università di Parma, Parma; ⁸Registro Tumori della Romagna, Ospedale GB Morgagni, Forlì; ⁹Registro Tumori della Provincia di Modena, Centro Oncologico Modenese, Azienda Ospedaliera Policlinico di Modena, Modena; ¹⁰Registro Tumori della Provincia di Ferrara, Università di Ferrara, Ferrara; ¹¹Registro Tumori di Popolazione della Regione Campania, c/o ASL NA4, Naples; ¹²Registro Tumori Piemonte, Provincia di Biella, Biella; ¹³Registro Tumori del Veneto, Istituto Oncologico Veneto, Università degli studi di Padova, Padova; ¹⁴Registro Tumori Reggiano, Azienda USL di Reggio Emilia, Reggio Emilia; ¹⁵Registro Tumori della Provincia di Trento, Azienda provinciale per i servizi sanitari, Trento; ¹⁶Registro Tumori Piemonte, Città di Torino, Turin; ¹⁷Registri Tumori della Provincia di Sassari ASL1-Sardegna, Sassari; ¹⁸Registro Tumori Lombardia, Fondazione IRCCS "Istituto Nazionale dei Tumori, Milan; ¹⁹UOS Registro Tumori della Provincia di Ragusa, Azienda ospedaliera "MP Arezzo", Ragusa; ²⁰Registro Tumori della Regione Liguria, SS Epidemiologia Descrittiva IST, DISSAL, Università di Genova, Genoa; ²¹Registro Tumori della Provincia di Macerata, Università di Camerino, Camerino (Macerata), Italy

Aims and background: The study aimed to validate model-based incidence estimates by means of observed incidence rates provided by Italian cancer registries, for five major cancer sites (stomach, colon and rectum, lung, breast and prostate cancers) and for all cancers together.

Methods: Recent incidence rates observed by Italian population-based cancer registries were extracted from the data base of the Italian Association of Cancer Registries. Regional estimates of incidence rates for the same cancers were obtained by the MIAMOD method. Observed and estimated crude incidence rates and incidence trends were compared for the period of diagnosis 1985-2000. Eight Italian cancer registries and seven regions were selected for the analysis since they had incidence data available during the entire selected period.

Key words: cancer incidence, Italy, registries, validation.

Introduction

The estimation of cancer incidence and prevalence carried out by the MIAMOD method^{1,2} is based on official mortality data and on relative survival observed by cancer registries (CRs). We applied the MIAMOD method to provide estimates of cancer incidence and prevalence at the national and regional level for the years 1970-1999 and projections up to the year 2010.

An alternative source of incidence data is given in Italy by population-based CRs. They collect very reliable observational data that, however, are subject to several limitations that prevent the possibility to provide a systematic and detailed picture of cancer burden during

Results and conclusions: An excellent agreement between estimated and observed crude incidence rates was found for all single cancer sites, regarding absolute incidence levels and time trends. A partial exception was breast, where empirical data showed a sudden increase in the last three years of observation, perhaps due to organized screenings in some Italian regions, and not captured by statistical models. Substantial underestimation of model-based incidence rates was found for all cancers combined, where the difference tended to increase with calendar year, up to a maximum of 20% in recent years. The greatest part of the discrepancy can be attributed to multiple cancers, which were included in cancer registries statistics but were not accounted for in MIAMOD estimates.

the considered decades. The Italian CRs cover only a subset of the national population, with a geographic distribution mainly centered in the northern regions. Not all the regions have at least one CR. Local CRs are not necessarily representative of the entire region. They cover sometimes a small population, and age- and time-specific incidence rates have wide confidence intervals. Furthermore, the period of registration differs for the various registries. Only three registries were active at the beginning of the 1980s, whereas 22 are active today with coverage of almost 20% of the Italian population.

However, data from CR are also an important benchmark for validating MIAMOD estimates, provided an appropriate strategy is designed to make the two sets of

Acknowledgments: The study was supported by: Foundation "Compagnia di San Paolo" (project "Sperimentazione di un sistema di rilevazione e diffusione dei dati di sopravvivenza per tumore"); Italian Ministry of Health (project "I TUMORI IN ITALIA" promoted by Alleanza Contro il Cancro).

Correspondence to: Dr Riccardo Capocaccia, Reparto di Epidemiologia dei Tumori, Centro Nazionale di Epidemiologia, Sorveglianza e Promozione della Salute, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy. Tel +39-06-49904290; fax +39-06-49904285; e-mail riccardo.capocaccia@iss.it

Received December 20, 2006; accepted March 5, 2007.

data as comparable as possible. The aim of the study was to validate MIAMOD-based estimates with CR-observed incidence rates for the five major cancer sites (stomach, colorectal, lung, breast and prostate cancers) and for all cancers together. Since a region-by-region direct validation is not possible or might be unreliable, we first restricted the comparison to the regions with at least one local CR. We then pooled the data for all the selected regions to remove random fluctuations. Furthermore, we chose to compare incidence trends instead of single-point estimates, since a correct estimation of time trends is a key point to obtain reliable incidence projections for the next years. Therefore, estimated and observed incidence trends for the period of diagnosis 1985-2000 were calculated. The validity of MIAMOD estimates and interpretation of differences between model-based and empirical-based incidence indicators is discussed.

Material and methods

We obtained the most updated incidence rates observed by 22 Italian population-based CRs extracted from the centralized data base of the Italian Association of Cancer Registries. The completeness of registration and the quality of the data of each CR were tested and ascertained before their inclusion in the data base³. MIAMOD-based estimated incidence rates, by cancer site, sex, age, year and region, are described elsewhere⁴⁻⁹ and are included in a data base available on the website www.tumori.net. Empirical and estimated incidence rates were both expressed as crude rates for the ages 0-84 years. The 0-99 year age interval was used only for prostate cancer.

We restricted our incidence trend analysis to the period 1985-2000 to maximize the inclusion of eligible CRs. Five Italian CRs had incidence data available during the entire period: Torino (the city), Parma, Romagna, Firenze and Ragusa CRs. In addition, we selected three other registries (Varese, Genoa [the city] and Veneto CRs) with missing data for at most two years at the start or at the end of the defined period. For the specific purposes of the analysis, incidence rates in the missing years were imputed by linear extrapolation. In synthesis, we obtained empirical incidence rates from eight registries, in seven Italian regions: Piemonte, Lombardia, Liguria, Veneto, Emilia Romagna (with two registries), Toscana and Sicilia.

A first average incidence trend indicator, named below as "pool" trend, was calculated by pooling together the crude age 0-84 incidence rates of the eight selected CRs, weighted by the corresponding person-years at risk in each registry. This indicator is the most intuitive one: it gives the lowest standard errors, but is scarcely sensitive to the incidence levels observed in CRs covering small populations, in particular in Ragusa CR, the only selected registry located in the South of Italy. Therefore, we calculated a second indicator, weighting age 0-84 crude incidence rates of each local CR by the corresponding regional population. This second trend indicator, named as "regional" trend, provides the incidence trend of the sev-

en selected regions, assuming that each CR can be considered as representative of the corresponding region. The "pool" and "regional" trends give, in general, different estimates. The MIAMOD estimated trends for the seven selected regions were pooled together giving the "MIAMOD" indicator. This corresponds to the "regional" indicator and provides the estimated trend in the selected regions.

Results

Figure 1 shows a comparison of the MIAMOD trend estimates and both the pool and regional (observed) crude incidence trends for each site/sex combination. The regional rates were systematically lower than the pool rates. This is mainly due to the contribution of the low incidence of the Ragusa CR, which is weighted much more in the regional trend, by the Sicily region population, than in the Pool trend, by the small Ragusa province population. MIAMOD estimates were generally closer to the regional than to the pool observed trend. The pool and regional crude incidence rates presented the same general shape and parallel random fluctuations, since they were based on the same empirical incidence rates.

Decreasing trends were shown for stomach cancer, in both sexes. MIAMOD crude incidence trends were very close, particularly in the last years, to the regional trend. Conversely, increasing trends were observed for colorectal cancer for both sexes. All the three approaches had the same linear slopes, and the MIAMOD estimate lay in between the two empirical trends. Male lung cancer crude incidence has been decreasing since the mid 1990s, whereas it has been linearly increasing for women. MIAMOD lung cancer estimates match well with the regional estimates. The crude incidence of breast cancer steadily increased until the year 1998, when both empirical approaches showed a peak in the last three years of observation. MIAMOD estimates matched perfectly with the regional method until 1998 but did not capture the final abrupt peak. The crude incidence of prostate cancer was also increasing, with a particularly marked rate of increase after the year 1990. The MIAMOD trend estimates match very well with the regional trends.

All cancers combined was the only case showing lack of agreement between MIAMOD estimates and empirical pool and regional trends. Pool and regional approaches showed increasing trends for both sexes. MIAMOD estimated rates were lower, and they increased with a lower rate than pool and regional approaches. MIAMOD crude incidence trend no longer increased for men and remained steady during the last five years. The discrepancy between trends for all cancers combined tends to increase with calendar year, up to a 20% difference in recent years.

A synthetic comparison between MIAMOD estimated and pool observed rates with respect to regional observed rates is presented in Table 1, using the percentage

Table 1 - Differences of MIAMOD incidence estimates and Pool observed incidence with respect to regional weighted observed incidence. Average percent differences over the period 1985-2000

Cancer site	Men		Women	
	MIAMOD (%)	Pool (%)	MIAMOD (%)	Pool (%)
Stomach	-2.6	+8.7	-1.6	+8.4
Colon & Rectum	+6.4	+10.2	+2.7	+9.5
Lung	+1.5	+10.2	-3.6	+18.0
Breast			-0.4	+7.0
Prostate	-0.2	+8.0		
All cancers	-10.0	+10.3	-11.1	+9.4

average difference of incidence rates over the period 1985-2000 as an indicator. For all single cancer sites, MIAMOD estimates were close to regional estimated rates, with average percentage differences ranging from -3.6% (lung cancer, women) to +6.4% (colorectal cancer, men). Pool incidence estimated rates showed larger differences, all with positive values indicating systematic overestimation. For all cancers together, the average differences from regional rates were about -10% for MIAMOD estimates and about +10% for pool rates, as visually shown by Figure 1.

Discussion

In this study, we validated the regional incidence estimates, obtained from mortality and survival rates by means of the MIAMOD model, using the corresponding observed incidence data provided by Italian CRs. The main advantage of the MIAMOD approach, with respect to other methods used for incidence estimation¹⁰, is that it allows time-trend estimates, including forward projections of incidence, prevalence and mortality rates. The availability of short-term projections referring to the current and to the next calendar years is particularly useful for health care planning purposes. Reliability of projections is however very sensitive to the validity of the assumed incidence rate of change. This is the reason why we focused the analysis on the validation of estimated incidence trends rather than on single point estimates.

A subset of eight CRs located in seven Italian regions was selected and pooled together in order to obtain robust empirical estimates of incidence trends based on a sufficiently long-term series and on the largest possible data set. CRs are not randomly distributed in Italy from a geographical point of view. They tend to concentrate in northern areas and are characterized by high incidence and survival rates. Therefore, the selected registries cannot be considered as a national sample, and we restricted the comparisons to the seven involved regions. The pool trends, where CR incidence rates were weighted by the size of the covered population, lie above the regional trends, in which each registry is weighted by the corresponding regional population, for all the site-sex combinations analyzed. MIAMOD-estimated trends are much closer to and in many cases practically overlap the regional trends.

For breast cancer, MIAMOD estimates fit very well regional trends up to the year 1998, when steep increases (+6% in 1998 and +7% in 1999) were reported by CR but were not captured by model-based estimates. The jump of observed rates was mainly due to the data from Torino CR (+12% in 1998 and +17% in 1999) and Varese CR (+16% in 1998). Sudden increases like those reported in the present paper have been previously observed¹¹ in other European countries, such as Sweden, Finland, and the UK, in correspondence with the start of breast cancer screening programs. We do not have sufficiently detailed data on screening activity to evaluate its impact on incidence trends in Italian regions, although we know¹² that several screening projects began in those years in Piemonte and Lombardia. In any case, the MIAMOD model is hardly able to capture changes in incidence rates that are not immediately reflected in mortality rates and in improvements of short-term survival rates.

For all cancers combined, the MIAMOD trend substantially underestimated observed regional trends. The differences were small in the first years but subsequently increased up to 20% for men and 15% for women. An important difference between MIAMOD incidence rates and CR reported incidence rates is that the former did not include second and subsequent primaries of the same type as the index tumor. This difference is practically appreciable only in the analysis of all cancers combined. Exclusion of multiple tumors from observed incidence is not always possible for registries with a not very long period of activity, because patients diagnosed with a second or subsequent cancer may be registered as first primaries just because their previous cancers were diagnosed before the start of registry activity. Anyway, after removing known cases of second or subsequent cancers, observed incidence rates decreased up to 10% in men and 7% in women in the year 2000, thus substantially reducing the gap between estimated and observed trends.

Statistical models are designed to provide correct and not systematically biased incidence estimates. However, they provide a necessarily simplified description of the reality, and incidence point estimates ranging within $\pm 10\%$ of the true value are usually considered as a satisfactory level of accuracy. In this paper, comparisons between estimated and observed crude incidence trends have shown a much better level of agreement, both in absolute level and in time changes, for most of the considered cancer sites and for both sexes. This indicates that MIAMOD estimates provide on the average correct estimates of incidence trends in Italy, but not necessarily that the same reliability can be expected for each single region estimates. Many different factors, including validity of the assumed survival model, proportion of the regional population covered by cancer registration, differences in registration techniques, and also statistical uncertainty can affect the validity of local estimates. However, apart from the problem of multiple tumors which practically affect only the estimates for all cancers combined, no systematic bias of MIAMOD estimates is expected even at the regional level.

Figure 1 - Comparison between observed incidence in eight Italian cancer registries (Torino, Varese, Genova, Veneto, Parma, Romagna, Firenze, and Ragusa) weighted by the registry (Observed Pool) and the regional population estimated incidence (Observed Regional) and MIAMOD estimated incidence (Estimated MIAMOD) in the same regions for stomach, colon and rectum, lung, prostate, female breast cancers and all cancers combined. Crude rates per 100,000, age 0-84 years (all ages for prostate cancer).

References

1. Verdecchia A, Capocaccia R, Egidi V, Golini A: A method for the estimation of chronic disease morbidity and trends from mortality data. *Stat Med*, 8: 201-206, 1989.
2. Verdecchia A, De Angelis R, Francisci S, Grande E: Methodology for estimation of cancer incidence, survival and prevalence in Italian regions. *Tumori*, 93: 337-344, 2007.
3. AIRT Working Group: Italian cancer figures, Report 2006: Incidence, mortality and estimates. *Epidemiol Prev*, 30 (Suppl 2), 2006.
4. Grande E, Inghelmann R, Francisci S, Verdecchia A, Micheli A, Baili P, Capocaccia R, De Angelis R: Regional estimates of all cancer malignancies in Italy. *Tumori*, 93: 345-351, 2007.
5. Grande E, Inghelmann R, Francisci S, Verdecchia A, Micheli A, Baili P, Capocaccia R, De Angelis R: Regional estimates of colorectal cancer burden in Italy. *Tumori*, 93: 352-359, 2007.
6. Grande E, Inghelmann R, Francisci S, Verdecchia A, Micheli A, Baili P, Capocaccia R, De Angelis R: Regional estimates of breast cancer burden in Italy. *Tumori*, 93: 374-379, 2007.
7. Inghelmann R, Grande E, Francisci S, Verdecchia A, Micheli A, Baili P, Capocaccia R, De Angelis R: Regional estimates of stomach cancer burden in Italy. *Tumori*, 93: 367-373, 2007.
8. Inghelmann R, Grande E, Francisci S, Verdecchia A, Micheli A, Baili P, Capocaccia R, De Angelis R: Regional estimates of lung cancer burden in Italy. *Tumori*, 93: 360-366, 2007.
9. Inghelmann R, Grande E, Francisci S, Verdecchia A, Micheli A, Baili P, Gatta G, Capocaccia R, Valdagni R, De Angelis R: Regional estimates of prostate cancer burden in Italy. *Tumori*, 93: 380-386, 2007.
10. Rosso S, Spitale A, Balzi D, Franceschi S, Zanetti R: Stima dell'incidenza dei tumori nelle regioni italiane nel 2001. *Epidemiol Prev*, 28: 247-257, 2004.
11. Sant M, Francisci S, Capocaccia R, Verdecchia A, Allemani C, Berrino F: Time trends of breast cancer survival in Europe in relation to incidence and mortality. *Int J Cancer*, 119: 2417-2422, 2006.
12. Giorgi D, Giordano L, Ventura L, Puliti D, Piccinni P, Paci E: Lo screening mammografico in Italia: survey 2003-2004. In: *Osservatorio Nazionale Screening. Quarto Rapporto*, pp 16-27, 2005.