Cancer incidence estimation method: an Apulian experience

Anna M. Nannavecchia[a], Ivan Rashid[a], Francesco Cuccaro[b], Antonio Chieti[a], Danila Bruno[a], Maria G. Burgio Lo Monaco[a], Cinzia Tanzarella[a] and Lucia Bisceglia[a]

The Cancer Registry of Puglia (RTP) was instituted in 2008 as a regional population-based cancer registry. It consists of six sections (Foggia, Barletta-Andria-Tran, Bari, Brindisi, Lecce, and Taranto) and covers more than 4 000 000 inhabitants. At present, four of six sections have received accreditation by AIRTUM (53% of regional population). To point out possible regional geographic variability in cancer incidence and also to support health services planning, we developed an original estimation method to ensure a complete territorial coverage. Incidence data of the four accredited RTP sections for the shared incidence period 2006–2008, the 2001–2009 hospitalization regional data, and 2006–2009 mortality data were considered. To take into account specific health features of different provinces, we performed an estimate of cancer incidence rates of nonaccredited sections using a combination of accredited sections rates and a factor that combines mortality and hospitalization ratios available for all the sections. Finally, we validated the method and we applied it to estimate regional cancer rates as the population-weighted average of accredited sections and nonaccredited sections adjusted rates. The validation process shows that estimated rates are close to real incidence data. The most frequent neoplasms in Apulia are breast (direct standardized rates 96.8 per 100 000 inhabitants), colon–rectum (36.6), and thyroid cancer (25.3) in women and prostate (70.2), lung (68.4), and colon–rectum cancer (52.2) in men. This method could be useful to assess the cancer incidence when complete cancer registration data are not available, but hospitalization, mortality, and neighbouring incidence data are available. *European Journal of Cancer Prevention* **26:S153–S156 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc.**

European Journal of Cancer Prevention 2017, **26**:S153–S156

Keywords: Apulia, cancer registry, incidence estimation, regional coverage

[a]Cancer Registry of Apulia, Health Regional Agency of Apulia and [b]Cancer Registry of Apulia, Local Health Unit of Barletta-Andria-Trani, Bari, Italy

Correspondence to Anna M. Nannavecchia, MD, Health Regional Agency of Apulia, Via Gentile 52, 70126 Bari, Italy
Tel: +39 080 540 3521; e-mail: am.nannavecchia@arespuglia.it

Received 16 January 2017 **Revised** 12 April 2017

Introduction

Knowledge of the neoplastic phenomenon through the main indicators, such as incidence, mortality, survival, and prevalence, is one of the pillars on which to base a proper allocation of the health resources. Whereas cancer mortality is a routinely recorded indicator in the Italian territory and is available at the provincial level, the incidence of tumors can be provided in an accurate and complete way only through a population-based cancer registry (CR) and survival and prevalence can be estimated on the basis of incidence data and follow-up.

CRs are recognized as high-quality instruments as they focus on the accuracy, standardization, and completeness of the incidence data of all malignant neoplasms; the quality of data is confirmed by an iterating check that inspects the proportion of DCO cases, the proportion of cases with microscopic confirmation, the mortality to incidence ratio, and other more sophisticated indicators. In addition, in Italy, the Italian Network of Cancer Registries (AIRTUM) ensures the reliability and the comparability of the CRs' data through a formal accreditation process. A CR can submit a request for accreditation only when it has completed the registration of at least 3 years' incidence, and then it has to provide a questionnaire with information on the methodology of registration and cancer coding, health sources available, epidemiological context, and high and stable quality checks. A specific commission analyzes questionnaire, incidence data, and individual and aggregated quality checks and after a site visit at CR, the commission will issue an opinion on acceptance or rejection of the application for accreditation. As a national coverage of neoplastic recording is not mandatory as yet, some areas of the country are not covered by accredited CRs and in some cases complete cancer incidence data do not exist at all.

In the region Apulia, a complete collection of regional cancer incidence is not yet available. Although there is a regional population-based cancer registry that consists of six provincial sections [Foggia (FG), Barletta-Andria-Trani (BT), Bari (BA), Brindisi (BR), Lecce (LE), and Taranto (TA)] covering more than 4 000 000 inhabitants, to date, only four out of six sections have been accredited by AIRTUM, whereas BA and FG are completing 3 years of incidence required for application for accreditation; the four accredited sections (BR, BT, LE, TA) cover 53% of

the regional population. To estimate the overall regional cancer incidence (Buzzoni *et al.*, 2016), mortality data and/or administrative databases – hospital discharge data, pharmaceutical data, payment exemption for disease, etc. – could be used. However, this provides only an approximation in the estimation of the cancer incident cases (Morgan and Scott, 1972; Toniolo *et al.*, 1986; McBean *et al.*, 1994; Solin *et al.*, 1994; Leung *et al.*, 1999; Couris *et al.*, 2002; Penberthy *et al.*, 2003; Brackley *et al.*, 2006; Gold and Do, 2007). Several studies have reported some difficulties when hospitalization data were used to identify the incidence of older individuals as successfully as younger individuals; indeed, older patients may have more comorbidities that likely affect the decision for surgery: physicians can avoid surgery and then decrease the likelihood of hospitalization of these patients for reasons related to the cancer.

The aim of the current paper is to validate the estimation methodology and to provide the first estimates of the 2006–2008 period cancer incidence for the whole region Apulia using complete data from an accredited cancer registry and mortality and hospitalization data available for the entire region.

$$\text{Rate}_I\ \text{NAS}\ (i,j)=\sum_{k=1}^{K}\text{rate}_I\ \text{AS}\ (i,j,k)\times\left[\left(\frac{\text{rate}_O\ \text{NAS}\ (i,j,k)}{\text{rate}_O\ \text{AS}\ (i,j,k)}\right)\times\left(1-\frac{M}{I}\ (i,j,k)\right)+\left(\frac{\text{rate}_M\ \text{NAS}\ (i,j,k)}{\text{rate}_M\ \text{AS}\ (i,j,k)}\times\frac{M}{I}\ (i,j,k)\right)\right],$$

Materials and methods

Incidence data of the four accredited sections (AS) of CR for the period 2006–2008 represent our starting point. We have a longer time coverage regarding administrative data of hospitalization and mortality; for this study, we used the 2001–2009 hospitalization and the 2006–2009 mortality of the entire region. Although we suppose a general territorial similarity of cancer incidence throughout the provinces of Apulian region, we also know that there is specificity for some neoplastic sites in each area that can be related to specific risk factors and/or local health organization. To take into account these two opposing issues (general homogeneity and local specificity), we applied the AS cancer incidence rates to nonaccredited sections (NAS) and used hospital discharge and mortality data available for each of the six provinces to stress the local epidemiological features of each area. We constructed an adjustment factor as a weighted combination of hospitalization and mortality ratios.

First, the hospitalization component was improved by identifying the first cancer occurrence of each patient in the 2006–2009 period, using 2001–2005 as the prevalent period for wash-out, that is, we eliminated all patients who had a hospitalization for the same tumor in the years before 2006. Thus, each patient was counted only once for tumor irrespective of the hospital in which the admission occurred. We call this kind of hospitalization information as 'refined hospitalization'. To represent the local propensity toward hospitalization, we calculated the ratio between the age-specific hospitalization rates in NAS and in AS. We also calculated the mortality ratio between NAS and AS age-specific mortality rates. Both of these components have been distinguished by sex and site and have been jointed through a linear combination. To confer appropriate weight to the two components, we hypothesized that mortality was a better proxy of incidence for more lethal tumors and, in contrast, hospitalization was a better proxy for less lethal tumors. We used the mortality/incidence ratio (M/I) as a proxy of cancer lethality; in particular, we applied it to the mortality component and its one's complement to the hospitalization one by each age group. The *M*/*I* ratio has been obtained by real incidence and mortality of AS. In CRs' context – where the *M*/*I* ratio is a completeness indicator – this ratio is calculated without age stratification, but it is an overall ratio by sex and site; in this context, given the different function of the *M*/*I* ratio, it was necessary to perform a weighted stratification by age group.

where i is the site, j is the sex, k is the age group, rate_I is the incidence rate, rate_O is the hospitalization rate, and rate_m is the mortality rate.

Finally, we obtained incidence rates for NAS and we validated the estimates by comparing them with NAS real incidence data available only for 2006. To obtain the overall incidence of the Apulian region, we constructed a population-weighted average of AS real rates and NAS estimated rates.

$$\text{Rate}_I\ \text{Apulia}\ (i,j)=\sum_{m=1}^{M}\sum_{k=1}^{K}\text{rate}_I\ (i,j,m,k)\times\frac{\text{pop}\ (m,k)}{\text{pop}\ (k)},$$

where i is the site, j is the sex, k is the age group, and m is the CR section.

A validation of our method involves the estimations of each AS incidence rates using the other AS real incidence rates. To validate AS rates, we calculated some validation indicators as Pearson's coefficient of correlation (r), coefficient of residual mass (Loague and Green, 1991) and efficiency model coefficient (E) (Nash and Sutcliffe, 1970).

Table 1 Observed and estimated incidence: direct standardized rates of Barletta-Andria-Trani, men and women, 2006–2008

	Male		Female	
Sites	Observed	Estimated	Observed	Estimated
Head and neck	21.4	18.2	3.6	3.9
Stomach	17.2	15.8	9.5	10.9
Colon–rectum	48.8	46.2	34.6	34.1
Liver	34.3	32.7	13.2	12.1
Pancreas	9.8	9.4	5.6	5.7
Lung	65.1	59.1	11.7	10.3
Mesothelioma	1.1	0.5	0.5	0.2
Skin melanoma	10.9	12.6	8.7	10.4
Prostate	72.2	74.6	–	–
Testis	5.3	5.6	–	–
Breast	–	–	102.9	106.2
Cervix	–	–	5.7	5.9
Corpus uteri	–	–	17.4	18.7
Ovary	–	–	13.5	11.7
Kidney	9.4	13.2	4.4	5.6
Urinary bladder	41.4	42.7	4.9	6.3
Central nervous system	8.5	9.4	5.8	5.6
Thyroid	5.1	4.1	21.6	17.9
Hodgkin's lymphoma	3.9	3.5	3.1	3.1
Non-Hodgkin lymphoma	15.2	15.9	11.2	12.6
Multiple myeloma	9.9	7.7	7.2	6.6
Leukemia	15.5	19.2	7.9	8.3

Results

The validation process applied to BT section (AS) shows estimated incidence rates closer to the observed ones (Table 1). We found good indicators of agreement between rates; in particular, the correlation coefficient is equal to 0.998, the efficiency model coefficient is 0.995, and the coefficient of residual mass is − 0.0109 among female patients and 0.993, 0.115, and 0.987 among male patients. Moreover, NAS estimates are also reliable and close to the real incidence data available for 2006. They are also in agreement with the expected health frameworks for each area. For instance, we found a higher rate of liver cancer in Bari province; this evidence is well known because it is confirmed by the mortality rate and also by a higher incidence rate of the adjoining province BT, which was recently instituted partially from Bari. Moreover, Bari province – which includes the metropolitan area of Bari – shows excesses for some sites: skin melanoma, in women, and testis and liver, in men Another notable result was the mesothelioma rate in Bari province, which is the highest of the region; this result has been confirmed by several cohort studies about workers exposed to asbestos in a cement plant of Bari (Nannavecchia *et al.*, 2016; Coviello *et al.*, 2002). Apulian incidence rates do not show outliers compared with Italy and Southern-Italy rates (Table 2). Apulian digestive system tumor rates are aligned to Southern-Italy, whereas lung cancer, in men, are aligned to overall Italian rates; in women, in contrast, lung cancer rates are aligned to Southern-Italy rates. In addition, we found a very similar ranking of cancer between Apulia and Italy (I numeri del cancro in Italia, 2016); the first five items are prostate, lung, colon–rectum, urinary bladder and stomach in Italian men. Liver cancer in Apulia replaces head and neck cancer. In women, the ranking of cancer prevalence is breast, colon–rectum, thyroid, lung, and corpus uteri in Italian women and breast, colon–rectum, thyroid, corpus uteri, and lung in Apulian women. Lung cancer shifts to the subsequent position in Apulian women in comparison with the Italian women.

Discussion

Our estimate of the regional incidence rate of lung cancer in women is lower than the Italian rate; a probable reason

Table 2 Bari, Foggia, and Apulia estimated direct standardized rates using the new method, Italy direct standardized rates and Southern-Italy direct standardized rates (AIRTUM data)

	Male					Female				
Sites	Bari*	Foggia*	Apulia*	Italy	South Italy	Bari*	Foggia*	Apulia*	Italy	South Italy
Head and neck	18.6	23.2	20.5	23.3	20.4	4.6	3.8	4.3	5.7	4.5
Stomach	14.4	17.0	15.6	21.1	14.8	8.2	9.1	8.1	10.6	7.4
Colon–rectum	52.8	59.4	52.2	64.2	49.7	37.1	39.6	36.6	40.3	33.4
Liver	25.3	19.9	22.7	20.3	18.9	8.2	6.3	7.2	6.5	7
Pancreas	11.4	9.9	10.7	12.7	10.2	7.5	6.1	6.8	9.6	7.3
Lung	60.2	59.1	68.4	69.3	63.9	12.3	11.6	12.3	19	12.8
Mesothelioma	2.8	1.6	2.4	2.7	2	0.8	0.3	0.6	0.8	0.4
Skin melanoma	11.2	9.2	9.3	12	7.2	12.1	9.3	9.8	11.4	6.6
Prostate	68.8	74.3	70.2	91.2	60.8	–	–	–	–	–
Testis	8.9	5.9	7.4	6.5	5.9	–	–	–	–	–
Breast	–	–	–	–	–	94.9	92.2	96.8	112.2	91
Cervix	–	–	–	–	–	5.3	6.9	6.3	6.2	5.8
Corpus uteri	–	–	–	–	–	16.6	16.6	17.0	17.8	16.5
Ovary	–	–	–	–	–	10.3	11.4	11.6	11.5	10.8
Kidney	12.1	11.8	12.0	18.7	12.5	6.1	6.0	5.9	8	5.1
Urinary bladder	48.6	46.8	49.9	47.9	49.1	6.0	5.3	6.2	8.9	7.9
Central nervous system	8.6	10.5	9.5	8.3	7.7	6.2	6.6	6.2	6	5.7
Thyroid	8.1	6.5	7.6	7.9	7.5	23.9	20.8	25.3	22.2	24.6
Hodgkin's lymphoma	2.6	3.5	3.4	3.9	3.5	3.3	3.4	3.6	3.4	3.3
Non-Hodgkin lymphoma	17.6	16.0	15.7	17.4	14.3	12.9	10.0	11.2	12.5	9.8
Multiple myeloma	6.8	6.4	6.5	6.3	4.8	5.5	4.6	5.6	4.5	3.9
Leukemia	15.1	16.7	15.5	12.4	12.3	9.9	9.9	11.5	7.9	8.1
All tumors but skin carcinomas	420.2	419.3	427.2	473.9	391.1	310.3	315.8	312.9	346.1	292.7

*Estimated.

could be related to female smoking attitude that concerns South Italy less and later in time than North; among men, the Apulian rate of lung cancer is equivalent to the Italian one. The skin melanoma rate is higher than the South Italian rate, but close to the Italian rate, especially because of Bari province, where the highest regional rate is recorded. The incidence rate of thyroid cancer is higher than the Italian and Southern rates – in women – because we suppose a larger local opportunistic propensity. Rates for liver cancer are higher than Italian and Southern rates, especially for the contribution of BT and Bari provinces; we are heavily dependent on studying the relation between this tumor and potential infectious factors such as hepatitis C virus infection and hepatitis and its different geographical distribution. Testis cancer incidence is higher in comparison with Italy and South Italy rate, and we found the highest rate in Bari province. The incidence of other tumors is aligned with South Italy rates, expect for screening cancer sites. This method provides reliable and likely cancer estimates; it could be useful to assess the cancer incidence when cancer registration data are not available for an area surrounded by areas with a CR. In our case, the aim has been achieved: we have finally estimated cancer rates for the entire region. This analysis provides a framework for public health planners to identify interventions and improvements around high-risk areas. Mortality data provide a good approximation of cancer incidence in cases of highly lethal cancers, but it is not a good estimator in other situations. Hospitalization alone is just a gross proxy of incidence. In fact, administrative databases lack the epidemiological purposes as a priority; they were originally intended for the refund rather than providing information on patients' health status. For example, in hospitalization data, the diagnosis is often not precise because the coding system (ICD-9th) is not accurate for describing neoplasm. In fact, a limitation of this method concerns the impossibility to identify tumor size and stage using hospitalization data only, whereas, as is known, CRs record all information to evaluate the incidence, including the stage. Our method, which uses real incidence data from neighbouring CRs and a combination of mortality and 'refined' hospitalization as an adjustment coefficient, seems to be able to estimate cancer phenomena where CR does not exist and to describe the cancer risk variability where CRs do not have full coverage of the area, but administrative data are available. The validation process confirms the goodness of the model; therefore, our next step could be an estimation of incidence trend.

Acknowledgements

Conflicts of interest

There are no conflicts of interest.

References

Brackley ME, Penning MJ, Lesperance ML (2006). In the absence of cancer registry data, is it sensible to assess incidence using hospital separation records? *Int J Equity Health* **5**:12.

Buzzoni C, Crocetti E, Airtum Working Group (2016). Stima di incidenza dei tumori nelle regioni italiane. 20th AIRTUM Congress book of abstracts. Available at: *http://www.registri-tumori.it/PDF/AIRTUM2016/PRESENTAZIONI/15-04-2016VEN/V_sessione/3_CarlottaBuzzoni.pdf*.

Couris CM, Colin C, Rabilloud M, Schott AM, Ecochard R (2002). Method of correction to assess the number of hospitalized incident breast cancer cases based on claims databases. *J Clin Epidemiol* **55**:386–391.

Coviello V, Carbonara M, Bisceglia L, Di Pierri C, Ferri GM, Lo Izzo A, *et al.* Mortalità di una coorte di lavoratori del cemento amianto a Bari., *Epidemiol Prev* 2002; **26**:65–70.

Gold HT, Do HT (2007). Evaluation of three algorithms to identify incident breast cancer in Medicare claims data. *Health Serv Res* **42**:2056–2069.

I numeri del cancro in Italia (2016). Airtum-Aiom. Available at: *http://www.registri-tumori.it/PDF/AIOM2016/I_numeri_del_cancro_2016.pdf*.

Leung KM, Hasan AG, Rees KS, Parker RG, Legorreta AP (1999). Patients with newly diagnosed carcinoma of the breast: validation of a claim-based identification algorithm. *J Clin Epidemiol* **52**:57–64.

Loague K, Green RW (1991). Statistical and graphical methods for evaluating solute transport models: overview and application. *J Contam Hydrol* **7**:51–73.

McBean AM, Warren JL, Babish Daniel (1994). Measuring the Incidence of cancer in elderly Americans using medicare claims data. *Cancer* **73**:2417–2425.

Morgan RW, Scott AE (1972). Hospital separations and cancer registration in British Columbia. *Can J Public Health* **63**:363–365.

Nannavecchia A, Cuccaro F, Bisceglia L, Coviello E, Baldassare A, Caputo E, *et al.* (2016). Mortalità in una coorte di lavoratori del cemento-amianto a Bari: aggiornamento al 2012, [Mortality in a cohort of asbestos cement workers in Bari, Italy: update to 2012]. Available at: *http://www.epicentro.iss.it/ben/2016/luglio-agosto/1.asp*.

Nash JE, Sutcliffe JV (1970). River flow forecasting through conceptual models part I – a discussion of principles. *J Hydrol* **10**:282–290.

Penberthy L, McClish D, Pugh A, Smith W, Manning C, Retchin S (2003). Using hospital discharge files to enhance cancer surveillance. *Am J Epidemiol* **158**:27–34.

Solin LJ, Legorreta A, Schultz DJ, Levin HA, Zatz S, Goodman RL (1994). Analysis of a claims database for the identification of patients with carcinoma of the breast. *J Med Syst* **18**:23–32.

Toniolo P, Pisani P, Vigano C, Gatta G, Repetto F (1986). Estimating incidence of cancer from a hospital discharge reporting system. *Rev Epidemiol Sante Publique* **34**:23–30.