Health Informatics Journal

Original Article

Health Informatics Journal
2019, Vol. 25(1) 149–160

DOI: 10.1177/1460458217704253
journals.sagepub.com/home/jhi
SAGE

Design, functionality, and validity of the SWInCaRe, a web-based application used to administer cancer registry records

Giovanni Benedetto, Alessia Di Prima, Salvatore Sciacca and Giuseppe Grosso

Integrated Cancer Registry of Catania-Messina-Siracusa-Enna, Azienda Ospedaliero-Universitaria "Policlinico-Vittorio Emanuele", Catania, Italy

Abstract

We described the design of a web-based application (the Software Integrated Cancer Registry—SWInCaRe) used to administer data in a cancer registry and tested its validity and usability. A sample of 11,680 records was considered to compare the manual and automatic procedures. Sensibility and specificity, the Health IT Usability Evaluation Scale, and a cost-efficiency analysis were tested. Several data sources were used to build data packages through text-mining and record linkage algorithms. The automatic procedure showed small yet measurable improvements in both data linkage process and cancer cases estimation. Users perceived the application as useful to improve the time of coding and difficulty of the process: both time and cost-analysis were in favor of the automatic procedure. The web-based application resulted in a useful tool for the cancer registry, but some improvements are necessary to overcome limitations observed and to further automatize the process.

Keywords

cancer case coding, cancer registry, record linkage, software, text-mining

Introduction

Cancer is among the leading cause of mortality affecting almost 15 million individuals and accounting for more than 8 million deaths worldwide.[1] In this scenario, information systems are needed to evaluate epidemiological parameters on cancer at population level, as well as to collect more detailed data on patients' demographic characteristics and clinical parameters potentially useful for in-depth studies on the relevant topic. Cancer registries have been established in several regions to collect information about new cases of cancer and to produce statistics about incidence, prevalence, survival, and mortality.[2] The process of identification and coding of cancer cases represents

Corresponding author:
Giuseppe Grosso, Integrated Cancer Registry of Catania-Messina-Siracusa-Enna, Azienda Ospedaliero-Universitaria "Policlinico-Vittorio Emanuele," Via S. Sofia 85, 95123 Catania, Italy.
Email: giuseppe.grosso@studium.unict.it

the main challenge to establish a cancer registry. Manual identification and coding is time consuming, money costing, and accuracy and validity of the process cannot be guaranteed.

Record linkage is a widely used process to link records derived by separate databases.[3,4] With the diffusion of large electronic health databases, the requirement for automated systems of record linkage has increased dramatically over the last decades.[5] Moreover, from a technical and financial point of view, the cross-link of multiple sources makes almost impossible the use of human resources to manually work on such prohibitively large data. Record linkage is a key component of cancer registries because case identification depends on the integrated information from various sources. The territorial distribution of the population requires a decentralization of the operators that triggers critical issues concerning the synchronization of the data collected and processed in the different areas. Moreover, the security management is challenged by the distribution of the databases through the territory and by the need to transmit data to the main servers. Once the cancer cases are identified, data must be coded according to the international coding rules before being analyzed and compared with other cancer registries data. Procedures of record linkage, tests for quality checks, and storage and control procedures of cancer cases are commonly used.[6,7] Software applications used for the management of data entry exist and their use has been widely validated.[8] However, record linkage associated with semi-automatic and automatic processing of incidence cancer cases strongly depend on the data source and ad hoc software build to optimize the workflow are highly demanded. Thus, the aim of this study was to describe methodology adopted to set up a web-based platform used in a regional cancer registry in Italy to collect and administer data on cancer cases and to describe its functionalities.

Methods

The following factors were considered to achieve a comprehensive description of the web-based platform:

- Design and functionality;
- Data quality (comparison of human (manual) *versus* software (automatic) procedure);
- Software usability;
- Cost-efficiency analysis;
- Random observation of system stability.

Setting

The cancer registry of Catania-Messina-Siracusa-Enna covers a population of approximately 2,300,000 inhabitants distributed across four main cities and a number of minor towns in the Eastern area of Sicily, Italy. The area involves a total of 207 municipalities and 79 main public hospitals.

Data sources

Data sources were the following:

- *Patient identification fiscal code* (FC), an univocal code determined through an algorithm using a person's name and his or her date and place of birth;
- *Regional identification registry* ("Nuova Anagrafica Regionale"—NAR), a regional database of all registered individuals living in the region (Sicily);

- *Hospital discharge record* (HDR), a form with an identification and information on type of procedures performed (including primary disease) provided at the moment of discharge from any Italian hospital. Only oncologic HDR are provided to the cancer registry, representing the main source of information for all cancers. All cases coded using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) are provided to the cancer registry;
- *Pathology record* (PR), available from suspect malignancy that undergoes biopsy or surgery to remove cells or tissues for examination under a microscope. The PRs provide necessary information such as topography and morphology as well as additional information, including the TNM stadium, tumor size, margins, vascular invasion, and lymph node status. These records allow to include patients who were assessed or treated in both hospital and outpatient setting (eventually not admitted at hospital nor diagnosed in primary hospitalization, such as screening program patients);
- *Mortality registry* ("Registri Nominativi delle Cause di Morte"—RENCAM), which are databases managed by the Health Local Authority keeping track of mortality of the population through and providing information on the cause of death according to the ICD-9;
- *Disease-specific exemption database*, which collect information on patients diagnosed of cancer by a specialist (oncologist) and provided documentation of their condition to Health Local Authority in order to obtain universal health insurance for medicaments and procedures;
- *Medical record* (MR), exceptionally required (case-sensitive) when no other source is available and is necessary to identify information relative to the case (either regarding clinical information or date of occurrence);
- *Other sources* include pharmaceutical prescriptions (File F) and diagnostic procedures (File P) related to oncological diseases.

Hardware and software characteristics

The web-based application that we created (the Software Integrated Cancer Registry—SWInCaRe) allows the storage of all oncological information provided either via a direct connection to the hospital servers or by data entry of individual information provided by the Regional Epidemiological Department in digital form. An SQL-server database and a programming language (Asp Dot Net) were used to store the data. To ensure the safety of access, we introduced 128-bit identification keys for the operators. Identification keys are renewed periodically (annually). According to the regulations for sensitive data in the health sector, HL7 languages are used for the connection to the main servers of the hospital associated with the cancer registry.

As information often relies on scanned paper/pdf files, text-mining algorithms were used if the available input databases were contained in text fields. Algorithms were designed in SQL-server. The procedure of text recognition was standardized first by definition of keywords in the database and then extraction of the information of interest through string searches.

Data of interest

According to the International Agency for Research on Cancer (IARC) guidelines, the basic information necessary to register a cancer case includes date of incidence, topography (anatomic site), and morphology. Anatomic codes are determined by the cancer sites origin, while the morphological codes are determined by the tissues and cells characterizing the cancer type, the levels of cancer differentiation, and the behaviors of tumor biology. Further European initiatives

(i.e. high-resolution studies of the EUROCARE project) require additional detailed information from MRs on representative samples of population-based cancer cases in order to conduct studies on clinical features of cancer patients.

Study sample

To test the quality of data collected, manual and automatic procedures were compared. A sample data from the province of Catania registered between 2003 and 2005 was considered for this study. Due to the high number of records, we selected only the surnames with the "C" letter (representing the most significant alphabetical group) including a total number of 11,680 records.

Human operators

The participating staff members consisted of six medical doctors who were provided the records to be examined and one information technology (IT) operator who worked on the web-based platform. The six medical doctors independently reviewed the records assigned to define incident, prevalent, and benign cases and resolved the existing discordances between the manual and the automatic procedures using the platform after a 1-week training; we considered this last procedure (automatic + manual check) as the gold standard to identify the best available estimate of the total number of cancer cases in this study.

Software usability

To evaluate the application usability, an Italian translation of the Health IT Usability Evaluation Scale was used.[9] The 20-item questionnaire is based on a 5-point Likert scale ranging from strongly disagree to strongly agree on four main domains: quality of work life, perceived usefulness, perceived ease to use, and user control. The score could range from 0 to 100, with score higher than 80 indicating high usability. The questionnaire was administered to the six medical doctors after use of the web-based platform.

Cost-efficiency analysis

Regarding the monetary comparison of the human resources needed for the operation compared with the software usage, we determined material and personnel costs in both cases. To assess the length of operations, completion time of the procedures (both manual and automatic) was taken from system log files. The costs of material acquisition and maintenance as well as personnel expenditures were collected through the administrative offices of the cancer registry. The calculation is based on the assumption that basic IT infrastructure, like printers and WiFi, already existed. Moreover, we did not include the costs for the operative system (such as various versions of Microsoft Windows) because it was used also for other tasks and in both procedures. To determine the personnel expenditures, we averaged hourly costs for healthcare/scientific assistant, taking into account a 25-Euro/h-threshold. We further calculated the cost-per-case by dividing the total costs for the number of cases found.

Statistical analysis

Continuous variables are presented as means and standard deviations (SDs) and categorical variables as frequencies and percentages. Specificity and sensibility were calculated for the manual

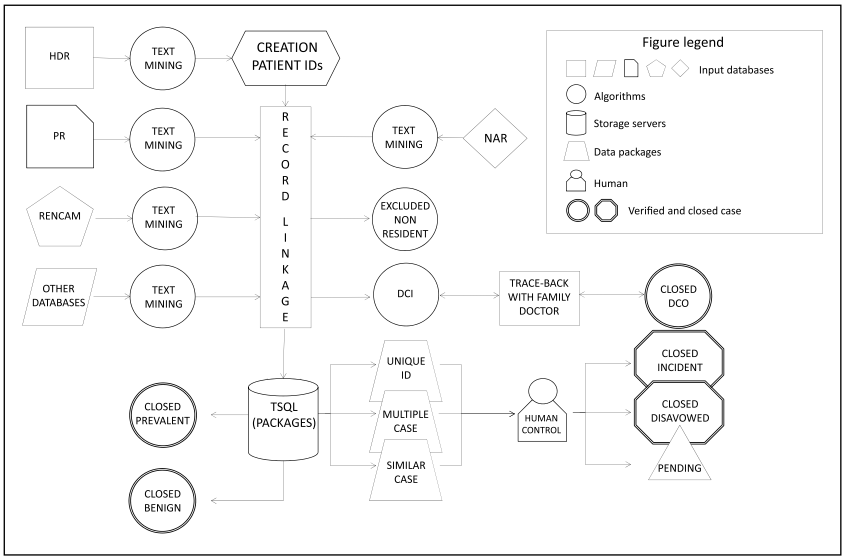

Figure 1. The Software Integrated Cancer Registry (SWInCaRe) implementation scheme.
DCI: death certificate initiated; DCO: death certificate only; HDR: hospital discharge record; NAR: regional identification registry; PR: pathology record; RENCAM: mortality registry; T-SQL: Transact–Structured Query Language.

and automatic procedures compared. All data were analyzed with Microsoft Excel (Microsoft, Washington, USA).

Results

Record linkage and case coding

The information collected is processed by the software according to the record linkage algorithms that allow to create potential univocal packages (Figure 1). A text-mining algorithm creates patients ID starting from the FC or from full name text, date of birth, and place of residence contained in data sources. Once the ID is created, all related information retrieved from any other available source related to the ID are matched to create univocal "data packages" stored in the database. A second algorithm searches for similarities between packages created through checking of identification components, such as name, surname, day, month, year of birth, and FC. The algorithm recognizes similarities as follows: (1) same name and surname without one letter or inverted, (2) same FC without last letter, and (3) inverted day and month of birth. If ID similarities refer to the same tumor, the algorithm merges the records; if ID similarities refer to different cancers potentially on the same patient, it creates a "multiple case"; if similarities are not sufficient to merge the records, it suggests the case as "similar," leaving to the human operator the choice whether to unify the records or not. IDs are automatically matched with the NAR and non-resident patients are excluded from the database automatically.

Besides the patient ID, the software automatically recognizes and provides a temporary topographic and morphologic diagnosis for all potential cases through the text-mining algorithm that recognize keywords from text of the HDR, PR, and RENCAM. The software encodes tumor anatomical codes automatically identifying from ICD-9. The morphological coding is relative to the biological behavior of the tumor, including malignant cancers, in situ cancers, benign, and

uncertain behavior cancers. Other morphological characteristics that the software recognize and register include the level of differentiation (such as undifferentiated, low, middle, and high differentiated), the types of tissues from which the tumor originated (such as epithelium, mesenchyme, lymph, hematopoietic, and nerve), and the type of cells from which the tumor originated (such as squamous, gland, basal, and transitional cells in the epithelial tissues).

The algorithm sets as primary potential data of incidence the oldest date provided by the HDR. Once the case is created, the operator can review all the information gathered in order to decide whether they are enough to close it. The case can be tagged as follows: (1) "verified incident," if diagnosis occurred over the period of observation; (2) "verified prevalent," if the diagnosis occurred previously, generally overlapping with previous closed cases; (3) "disavowed," whether the diagnosis was of benign cancer or not tumor disease. If the minimum dataset necessary to close a case is reached (including date of incidence, morphology, and topography), the case can be saved as verified and stored as incident; otherwise, the case can be saved as "pending" for further information.

Human versus software procedure

Six operators were asked to review the 11,680 records (including HDR, RENCAM, PR) in order to define incident, prevalent, and disavowed cases, including merging potentially overlapping cases, multiple cases (several cancer sites in the same patient), and check for actual residency in the area associated with the cancer registry during the period of diagnosis (Table 1). The process required 8 h and led to the identification of 4267 potential univocal packages. The record linkage was manually performed by linking the ID retrieved with all sources available for clinical information (HDR, PRs, etc.). The process took a total of 620 h for all operators and led to the identification of 2713 incident malignancies, 879 benign cases (disavowed), and 675 prevalent cases. In all, 152 were multiple cases. The same sample was tested with the automatic procedure of the software. The processing time was about 3 min. The system identified 4169 unique packages (Table 1). The record linkage with the sources and the creation of the cases took about 7 h. A total of 2696 malign incident cases, 869 benign, 604 prevalent cases, and 4 non-residents were excluded cases. The same number of multiple cases was retrieved.

Both procedures were reviewed by the six operators that manually validated the cases created, reducing to a total of 2561 incident cases (including 862 benign, 604 prevalent cases, and 4 non-residents). The discordant cases identified through the manual procedure were due to human error in identification of same patients (mistakes in univocal packages retrieved). The discordant cases identified through the automatic procedure were due to mismatch between data sources ID information (incorrect names and surnames, FCs, birth dates, derived by human data entry). However, part of such mistakes were identified by the algorithm as "similar cases" and left to the human review (data not shown).

Software usability

The graphical user interface was kept simple and consistent throughout the entire application. The landing page provides several search fields in order to allow to retrieve cases by name/surname, FC, and case characteristics (year of diagnosis, topography, morphology, mortality, and status (to be verified, verified, pending); Figure 2). Once entered a case, the application interface is designed to provide all mandatory information to close a case (according to IARC guidelines) always visible (Figure 2) and a further drop down menu to provide additional information required for high-definition studies (Figure 2). A list of all data collectable and respective source is shown in Table 2.

Table 1. Manual and automatic procedures to identify cases from a sample of 11,680 total records (including hospital discharge records, pathology records, and death certificates).

	Human (manual) procedure	Software (automatic) procedure	Confirmed (automatic + manual review)	Human procedure		Software procedure	
				Sensitivity	Specificity	Sensitivity	Specificity
Univocal packages, n (%)[a]	4267	4169	4185	94.5%	97.0%	96.6%	98.1%
Incident malignant cases, n (%)[b]	2713	2696	2561	92.7%	88.5%	93.3%	89.4%
Multiple incident cases, n (%)[b]	152	152	152	100%	100%	100%	100%
Incident benign cases, n (%)[b]	879	869	862	98.0%	99.4%	99.2%	99.8%
Prevalent cases, n (%)[b]	675	604	604	82.2%	94.7%	100%	100%
Non-resident patients, n (%)[b]	–	4	4	NA	NA	100%	100%

NA: not applicable.
[a]Total number refers to the total hospital discharge records examined.
[b]Total number refers to univocal packages retrieved.

Figure 2. Basic interface of the SWInCaRe.

Eye-catching icon-buttons also allow to request additional data sources (including MRs, pathology reports, or contact of general practitioner of the patient) in case needed.

The evaluation of the completed Health IT Usability Evaluation Scale questionnaires showed an overall score of 90 (2.6 SD) out of a maximum of 100, indicating high usability of the product. The single evaluation of all questions can be found in Table 3. Among those statements reaching higher scores, those mainly related to quality of work life showed complete agreement among users. In contrast, those items related to ease to use showed slightly lower scores, suggesting that the

Table 2. Variable domains collected through the SWInCaRe.

Variable domains	Input database	Notes
Personal data		
Sex, age, residence, date of birth	HDR, PR, File F, File P	Automatic
Identification codes		
Patient identification number		Auto-created
Tumor characteristics		
Incidence date, ICDO3 topography, ICDO3 morphology	HDR, PR, RENCAM	Automatic
Laterality, dimension, TNM Stage, grade, diagnosis source, positive lymph nodes, total lymph nodes analyzed	PR, MR (occasional), family doctor (occasional)	Manual
Gleason score, grade, resection margin (for prostate cancer)	PR	Automatic
Receptor status, hercept test, cerb2, vascular invasion, resection margin, sentinel lymph node (for breast cancer)	PR, MR (occasional), family doctor (occasional)	Manual
Clark, Breslow, sentinel lymph node (for melanoma cancer)	PR, MR (occasional), family doctor (occasional)	Manual
Dukes, Aslter Coller (for colorectal cancer)	PR, MR (occasional), family doctor (occasional)	Manual
Grade, resection margin (for skin cancer)	PR	Automatic
Chemotherapy, radiotherapy, surgery	HDR, MR, File F, File T	Automatic (except from clinical records)
Follow-up		
Date of follow-up	HDR, MR (occasional), family doctor (occasional), RENCAM	Automatic (except from clinical records)

SWInCaRe: Software Integrated Cancer Registry; File F: pharmaceutical prescriptions; File P: diagnostic procedures; HDR: hospital discharge records; MR: medical records; PR: pathology records; RENCAM: mortality registry.

learning curve may be longer than expected. Finally, the interviewed agreed that the web platform helped them more likely to code faster than more correctly.

Monetary analysis

A total of 620 h for six operators were needed for the identification of potential univocal packages (8 h) and potential cases (612 h) through the manual procedure: the total amount of pay-per-hour would correspond to 15,500 Euros and about 4 Euros per case. The automatic procedure required 8 h of work for one operator and two servers for data storage for a total of 250 Euros and about 5 Euro Cents per case.

System stability

We found no particular instabilities of the automatic system. However, being the system web-based, we observed some slowdowns in occasion of some troubles occurring at the central Internet connection provider, which caused long loading periods and, in some cases, even system crashes. In occurrence of session crash, information imputed by the operator were lost if not saved previously.

Table 3. Modified version of the Health IT Usability Evaluation Scale adapted to test the SWInCaRe usability.

Questions	Scores, mean (SD)
Quality of work life	
1. I think SWInCaRe has been a positive addition to coding procedures	5.0 (0.0)
2. I think SWInCaRe has been a positive addition to our organization	5.0 (0.0)
3. SWInCaRe is an important part of our coding process	5.0 (0.0)
Perceived usefulness	
1. Using SWInCaRe makes it easier to code	5.0 (0.0)
2. Using SWInCaRe enables me to code more quickly	5.0 (0.0)
3. Using SWInCaRe makes it more likely that I will code correctly	1.6 (0.8)
4. Using SWInCaRe is useful for coding	5.0 (0.0)
5. I think SWInCaRe presents a more equitable process for coding	5.0 (0.0)
6. I am satisfied with SWInCaRe for coding	4.3 (0.8)
7. I code in a timely manner because of SWInCaRe	5.0 (0.0)
8. Using SWInCaRe increases number of coded cases	5.0 (0.0)
9. I am able to access information to code whenever I use SWInCaRe	4.6 (0.5)
Perceived ease of use	
1. I am comfortable with my ability to use SWInCaRe	4.5 (0.5)
2. Learning to operate SWInCaRe is easy for me	4.5 (0.5)
3. It is easy for me to become skillful at using SWInCaRe	4.0 (0.0)
4. I find SWInCaRe easy to use	3.3 (0.5)
5. I can always remember how to log on to and use SWInCaRe	5.0 (0.0)
User control	
1. SWInCaRe gives error messages that clearly tell me how to fix problems	4.0 (0.0)
2. Whenever I make a mistake using SWInCaRe, I recover easily and quickly	3.3 (0.5)
3. The information (such as online help, on-screen messages, and other documentation) provided with SWInCaRe is clear	5.0 (0.0)
Total score	90.0 (2.6)

SWInCaRe: Software Integrated Cancer Registry; SD: standard deviation.

Discussion

The aim of this study was to describe the rationale behind the creation of a web-based platform able to administer data for a cancer registry and validate its functionality and usability. We reported an optimal performance of the algorithms as well as a good usability of the platform. Some limitations have, however, emerged and commented.

Linking HDRs with other registries data has emerged as a major source of gaining diagnosis and treatment procedure information related to cancer.[10] The process of designing and testing of the web platform required the work of several specialists providing individual expertise on the topic: epidemiologists provided support to design the functionalities needed to collect crucial information to identify the cases and to share the data with international bodies; clinicians, pathologists, and oncologists identified clinical core information to be added to minimum standard data collected; and IT experts, necessary to program the platform as well as to manage data input and databases.[11] We used external information through linkage with several data sources and algorithms able to identify univocal data packages (in most cases completion of topography and morphology) that can be validated by the human operator in order to identify and close a case.

The main advantage of using several data sources is that increase the algorithm capacity to identify the patients ID as well as to allow more complete and precise information recorded about the individuals of interest.[4] The main limitation for the application relied on local issues related to inaccuracy of the sources due to typo or transcription errors regarding the ID information. The errors encountered may regard incompleteness or omission of second names/surnames and errors of the FCs. Another limitation depended on the PRs, which are still manually registered (thus subjective of typo errors) and significantly missing of important information necessary for the identification of the patient and record linkage process (i.e. missing date and place of birth). However, errors were overcome by manual check of cases, leading to a proper functioning of the entire system with a minimum manual work on the automated procedure provided by the algorithms.

The automatic procedure showed small yet measurable improvements in both data linkage process and estimating cancer cases. However, the main goal of the application was to reduce the time of coding, rather than the quality itself. Commercial and ad hoc programmed software for data linkage are commonly used in cancer registries.[12,13] In epidemiological studies, false-positive linkages result in underestimation of true rates, whereas false-negative linkages result in overestimating rates. As small errors in record linkage (5%) can yield a significant error in estimating true rates, both procedures tested require further control by human operators.[14] However, the web-based application is designed to aid human operator to code cancer cases rather than automatically code and close the cases. The application usability has been tested showing high scores especially on the work quality and usefulness. Regarding the latter, users did not perceive the application as useful to improve the quality of coded data, despite our analysis showed a more accurate coding through the automatic procedure than the manual ones. However, the time analysis and the easiness to code were obviously in favor of the automatic procedure. A potential contributor to the usability may be the web-based solution, which did not require installation of the program and increased usability through access from any device.

The results of this study should be considered in light of some limitations. As mentioned before, the accuracy of the record linkage performed by the application depend on the quality of the information included in the database, which in our geographical area of application are subject to lack of digitalization, lack of barcode IDs, and typo mistakes. Moreover, the core of the information is based on the HDRs, which are administrative data collected to inform payment and billing operations, rather than clinical care. Thus, using them for clinical purposes requires some degree of inference and is yet not sufficient to provide full information for the minimum dataset as well as for additional inquiries (for instance, register cancer recurrence). Another limitation that should be taken into consideration with text-mining is that the algorithms are not univocal but must be continuously updated on the basis of the information that should be retrieved by the text, which can be presented in various ways if not included within a template form (for instance, pathologist can use different sentences when describing the variables extracted from the pathologic anatomy records). Lack of a unique person identifier within the country does not permit to overcome the problems related to record linkage and all data privacy concern.

In conclusions, the use of our web-based application on administrative databases to build patients registries offers great opportunities to enhance cancer-related research through the study of large numbers of patients. These resources can help us better understand cancer treatment outcomes, quality of care, resource utilization, and clinical management. Finally, further record linkage with other administrative databases on diseases that can be tracked (i.e. diabetes) would amplify the utility of software related to cancer registries and provide insights on cross-link between diseases at population level.

Acknowledgements

B.G. designed the software and wrote the manuscript, A.D.P. designed and programmed the software, S.S. conceived and provided insights on the methodology, and G.G. conceived and wrote the manuscript. All authors provided critical revision of the manuscript.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Regional Office of Health, Sicily, Italy.

References

1. Global Burden of Disease Cancer Collaboration, Fitzmaurice C, Dicker D, et al. The global burden of cancer 2013. *JAMA Oncol* 2015; 1: 505–527.
2. Cancer registration: principles and methods. *IARC Sci Publ* 1991; 95: 1–288.
3. Contiero P, Tittarelli A, Tagliabue G, et al. The EpiLink record linkage software: presentation and results of linkage test on cancer registry files. *Methods Inf Med* 2005; 44: 66–71.
4. Ryan R, Vernon S, Lawrence G, et al. Use of name recognition software, census data and multiple imputation to predict missing data on ethnicity: application to cancer registry records. *BMC Med Inform Decis Mak* 2012; 12: 3.
5. Wei KR, Liu SC, Wei D, et al. Auto-coding of cancer registry data in China. *Asian Pac J Cancer Prev* 2016; 17: 3021–3023.
6. Shats O, Goldner W, Feng J, et al. Thyroid cancer and rumor collaborative registry (TCCR). *Cancer Inform* 2016; 15: 73–79.
7. Sherman S, Shats O, Fleissner E, et al. Multicenter breast cancer collaborative registry. *Cancer Inform* 2011; 10: 217–226.
8. Campbell KM, Deck D and Krupski A. Record linkage software in the public domain: a comparison of Link Plus, The Link King, and a "basic" deterministic algorithm. *Health Informatics J* 2008; 14: 5–15.
9. Yen PY, Wantland D and Bakken S. Development of a customizable health IT usability evaluation scale. *AMIA Annu Symp Proc* 2010; 2010: 917–921.
10. Lin G, Ma J, Zhang L, et al. Linking cancer registry and hospital discharge data for treatment surveillance. *Health Informatics J* 2013; 19: 127–136.
11. Baili P, Torresani M, Agresti R, et al. A breast cancer clinical registry in an Italian comprehensive cancer center: an instrument for descriptive, clinical, and experimental research. *Tumori* 2015; 101: 440–446.
12. Marquez Cid M, Chirlaque MD and Navarro C. DataLink record linkage software applied to the cancer registry of Murcia, Spain. *Methods Inf Med* 2008; 47: 448–453.
13. Oberaigner W and Stuhlinger W. Record linkage in the Cancer Registry of Tyrol, Austria. *Methods Inf Med* 2005; 44: 626–630.
14. Gardner J, Xiong L, Xiao Y, et al. SHARE: system design and case studies for statistical health information release. *J Am Med Inform Assoc* 2013; 20: 109–116.