

## Data linkage and the South Australian Cancer Registry

Wayne Clapton<sup>1,2</sup>, Colin Luke<sup>3</sup>, Peter Chapman<sup>4</sup>, Kevin Priest<sup>5</sup>, Anh-Minh Thi Nguyen<sup>5</sup>, Graeme Tucker<sup>5</sup>, Maria Cirillo<sup>2</sup>, Mary Merdo<sup>2</sup>, Teresa Molik<sup>2</sup>, Elaine Morton<sup>2</sup>, Chris Scott<sup>2</sup>, Heather Hall<sup>2</sup>, Joanne Bell<sup>2</sup>, Lesley Milliken<sup>3</sup>

1Anti-Cancer Foundation of South Australia

2South Australian Cancer Registry (Epidemiology Branch)

3Clinical Epidemiology Unit (Epidemiology Branch)

4Epidemiology Branch

5Health Statistics Unit (Epidemiology Branch)

### Dr Wayne Clapton

Specialist Medical Officer (Public Health Physician), c/- SA Cancer Registry, Epidemiology Branch, Department of Human Services, PO Box 6, Rundle Mall, South Australia 5000, ph +61 (0)8 8226 6362, fax +61 (0)8 8226 6291, email Wayne.Clapton@dhs.sa.gov.au, web <http://www.dhs.sa.gov.au/pehs/disease-control-stats.htm>

#### Abstract

The South Australian Cancer Registry provides timely and accurate cancer epidemiological surveillance services for the State of South Australia. Cancer is a legally notifiable disease. Under specific State legislation, all hospitals, pathology laboratories and radiation oncology treatment centres are required to report all cases of invasive cancer (excluding non-melanotic skin cancers) to the Cancer Registry within one month of discharge and/or finalisation of diagnosis. The legislation also provides legal protection for those parties specified as exchanging data with the Cancer Registry. The Registry is required to protect the privacy and confidentiality of the data and not release it to any unauthorised party.

Cancer Registries aim to achieve 100 percent capture of cases (100% ascertainment), to minimise case duplication and to be aware of major events, such as the death of a patient. Various strategies are employed to this end, including correlating data from different sources, obtaining data from Births Deaths and Marriages, and linking Cancer Registry data with other databases such as hospital administration systems, national cancer data accumulations at the Australian Institute of Health and Welfare, and the National Death Index. Probabilistic matching processes are used mostly; but deterministic methods also may be employed.

This paper will summarise the importance of data linkages for the SA Cancer Registry, will suggest other currently unavailable linkages which would be useful, will discuss the importance of legislation in terms of its enabling and inhibitory effects on health data linkage and will postulate on the role of a Cancer Registry in a health system where data linkage is more prevalent.

#### Introduction

The South Australian Cancer Registry (SACR) provides timely, accurate cancer epidemiology surveillance services for the state of South Australia. The SACR started operations in 1977. The information it produces is used widely for activities in cancer prevention, control, treatment, education, research and health services planning. Data linkage is an integral feature of the work of the SACR. This paper initially will set the scene with some general points about the epidemiology of cancer. Then it will describe briefly the common features of a cancer registry and will provide a very limited overview of the South Australian Cancer Registry's activities. Following this will be a summary of the current use of data linkage in the SACR. Other potential uses of data linkage in the SACR then will be proposed and potential inhibitions of these activities will be postulated. A comment will be made of the proposed future role of a cancer registry in a "linked world".

#### Cancer epidemiology – In brief

Cancer is a common and serious disease. The 2001 South Australian Cancer Registry Report<sup>1</sup> provides figures for all cancers, (excluding non-melanotic skin cancers) in South Australia for the years 1977 to 2000. In 2000, for males, there were 4,126 new cancer cases diagnosed (average of about 11 per day). This corresponded to an age-standardised (to the world population) incidence rate of 351.0/100,000 males. There were 1,800 male deaths from cancer (average of about 5 per day), corresponding to an age-standardised (to the world population) death rate of 137.5/100,000 males. For females, there were 3,410 new cancer cases diagnosed (average of about 9 per day), corresponding to an age-standardised incidence rate of 270.8/100,000 females.

Female cancer deaths numbered 1,358 (average of about 4 per day), corresponding to an age-standardised death rate of 88.2/100,000 females.

The most recent cancer statistics for Australia as a whole are for 1998 and can be found in the Australian Institute of Health and Welfare (AIHW) & Australasian Association of Cancer Registries (AACR) "Cancer in Australia 1998" October 2001 report.<sup>2</sup> Again, these are figures for all cancers, excluding non-melanotic skin cancers.

For males, there were 43,595 new cancer cases diagnosed (average of about 119 per day). This corresponded to an age-standardised (to the world population) incidence rate of 340.9/100,000 males. There were 19,400 male deaths from cancer (average of about 53 per day), corresponding to an age-standardised (to the world population) death rate of 143.5/100,000 males. For females, there were 37,269 new cancer cases diagnosed (average of about 102 per day), corresponding to an age-standardised incidence rate of 268.7/100,000 females. Female cancer deaths numbered 14,870 (average of about 41 per day), corresponding to an age-standardised death rate of 91.9/100,000 females.

The same AIHW & AACR report indicates that 1 in every 3 males and 1 in every 4 females will contract cancer before the age of 75 years. Cancer is one of the "National Health Priority Areas", which include Cardiovascular Health, Cancer Control, Mental Health, Injury Prevention and Control, Diabetes Mellitus and Asthma (<http://www.health.gov.au/hsdd/nhpq/3rdnhpa.htm>). In terms of contribution to total burden of disease and Injury as measured by Disability Adjusted Life Year (DALY) scores, cancer ranks second as follows: Cardiovascular Disease 20%; Cancer 19% and Mental Disorders 14% (Mathers et al., 1999).<sup>3</sup>

### Common features of a Cancer Registry

A cancer registry is a repository of data on cancer in the community. An analogy might be an ongoing census. It collects data on ALL cases of cancer in a defined population. In all states of Australia, cancer is a legally mandated notifiable disease. A cancer registry is a public health monitoring and surveillance tool for assessing the epidemiology of cancer in the community. It is an applied public health practice activity, rather than being a research endeavour.

A cancer registry aims to collect information on ALL cases of cancer (that is, FULL ASCERTAINMENT) within a defined population. The data sources usually involve a combination of some or all of the following: hospitals, pathology laboratories, radiation oncology treatment centres, nursing homes, medical practitioners, coroners' offices and the registrar of births, deaths and marriages. Data from different sources are collated, interpreted, coded and recorded in the central registry and problems with identifying and defining cases and cancers are resolved.

The accumulated complete data enables the description of raw numbers of the various types of cancer and the calculation of INCIDENCE rates (new cases per 100,000 population in a

defined time period, usually one year) and MORTALITY rates (deaths from cancer per 100,000 population in the same time period). Variations of these rates over time can be monitored to determine the effects of various preventative, control and/or treatment interventions at the population level. Furthermore, as accurate data on times and causes of death are collected, SURVIVAL RATES can be calculated for various forms of cancer.

Incidence, mortality and survival rates usually are STANDARDISED by age and gender to some population standard (for example, the world population) to allow comparisons to be made between populations and time periods with different mixes of age and gender proportions. Estimates of PREVALENCE (number of existing cases in the population at a particular time point or over a time period) may be done as well; or this can be estimated separately by population sampling in surveys.

A good quality cancer registry provides accurate, complete, timely epidemiological information about the extent, burden and trends over time of cancer in the community. This enables informed, logical, sensible decisions to be made about activities in cancer prevention, control, treatment, health services planning, education and research.

An overview of the principles and methods of cancer registration can be found in Jensen et al (1991).<sup>4</sup> and Esteban et al (1995).<sup>5</sup>

### Limited overview of the SA Cancer Registry

#### Legal Underpinning

The SA Cancer Registry has a specific legal underpinning. Cancer is a legally mandated notifiable disease (without penalty) under the South Australian Health Commission (Cancer) Regulations 1991, No. 73 (<http://www.parliament.sa.gov.au:8080/ISYSquery/IRLC814.tmp/1/doc>). All hospitals, radiotherapy units and pathology laboratories are obliged to report all cases of cancer (except non-melanotic skin cancers) within one month of discharge and/or diagnosis.

Section 64d, SA Health Commission Act 1976, which became the South Australian Health Commission (Administrative Arrangements) Amendment Act 2000 No. 34 of 2000, specifies the conditions for the "disclosure of confidential information for certain purposes" (<http://www.parliament.sa.gov.au:8080/legsearch.htm>). This section authorises access to confidential information for the purpose of conducting research into the causes of mortality and morbidity; or for assessing and improving the quality of specified health services. Committees, institutions, and individuals are gazetted (eg, "Epidemiology Branch" is a legally gazetted name)

There is a prohibition on divulging the confidential information to which the person has access to anybody or any organisation not covered under Section 64d; including any court, tribunal or board. However, the confidential information may be disclosed by a person to whom this section applies to another person also covered under this section of the Act.

### Data capture, storage, processing, analyses and uses of outputs

Hospitals, radiation oncology units and pathology laboratories are required to report all cases of cancer to the cancer registry within a one month time-frame. Deaths are captured from the Registrar of Births Deaths and Marriages locally and the National Death Index (derived from Australian Bureau of Statistics data) in Canberra.

Resolution of data problems is performed via interactions with the data suppliers, interstate cancer registries and/or medical practitioners, who help voluntarily. The SACR has very good relationships with the local medical community and receives excellent cooperation, based on mutual respect and trust that data will not be mis-used. This is a very important “non-technical” factor for ensuring the successful operation of the SACR.

Following central cancer registry data collation, resolution of any problems and coding, data are entered onto a computer database from where de-identified data are extracted for analyses and reporting.

#### Analysed and reported data are used for:

- Assessing the burden of cancer, and of its various types, in the community at particular time points and over time (time series monitoring).
- Comparing different population groups and different geographical areas.
- Providing early warning of “cancer epidemics”.
- Evaluating the effects of interventions at a population level (eg, population-based screening programs).
- Contributing to research, (eg, suggesting hypotheses for aetiological or interventional studies).
- Contributing to health services planning activities.
- Contributing to disease prevention and health promotion programs.
- Supporting clinical services and quality assurance activities.

#### The following summarises some of the routine uses of SACR information output and interactions with other institutions and/or endeavours:

##### International

- International Agency for Research on Cancer (IARC), Lyon, France
- International Association of Cancer Registries (IACR)

##### National

- National Cancer Statistics Clearing House (NCSCH), Canberra
- Australian Institute of Health and Welfare (AIHW), Canberra

- Australasian Association of Cancer Registries (AACR)
- Other Australian Cancer Registries

##### State/Local

- Case notifications to Hospital-Based Cancer Registries
- BreastScreenSA (Interval cancer detection)
- SA Cervix Screening Program (Data quality assurance)
- Interactions with other cancer data collections
- Other government and non-government programs (eg, the Anti-Cancer Foundation)

#### The following summarises some of the ad hoc uses of SACR information output and interactions with other institutions and/or endeavours:

- National Cancer Survival project
- AIHW/AACR Ductal Carcinoma in situ (DCIS) Report
- National Colorectal Cancer Care Survey
- National Breast Cancer Care Survey
- Health Status Follow-up Studies
- Perceived Cancer Clusters in the community (usually in association with some environmental concern)
- Metropolitan Services Reviews
- National Public Health Partnerships activities
- Data for Policy Reviews
- Health services utilisation reviews
- National Cancer Control Initiative (NCCI) activities
- Australian Cancer Society activities
- Anti-Cancer Foundation of South Australia activities
- Australasian Association of Cancer Registries activities
- Health statistics chartbooks

Privacy, confidentiality and data security issues are taken very seriously and addressed comprehensively in the work of the SACR.

#### Current use of data linkage – SA Cancer Registry

##### Registrar of Births Deaths and Marriages (BDM) Death Linkages

Each month, the latest death certificates from BDM are reviewed by Cancer Registry Officers, who detect deaths due to cancer. Either they match that death certificate with a case on the Cancer Registry or initiate an investigation to capture that case if they are not on the Registry already. This is a manual process.

Each year, or preferably on 3–4 occasions per year, the accumulated previous year’s BDM death data is matched against the Cancer Registry database live South Australian cases. This

enables the detection of deaths for Cancer Registry cases who have died from causes other than cancer, or some deaths from cancer which may have been missed in the monthly processing. This is done by using a locally produced program or by using the "Automatch"<sup>6</sup> software.

#### **Australian Institute of Health and Welfare National Death Index Linkage**

On an annual basis, all South Australian live cases from 1977 to the most current and complete year (in terms of data entered) are matched against the National Death Index in Canberra, using the "Integrity" software, (which includes "Automatch"). The National Death Index contains the accumulated deaths as coded for cause by the Australian Bureau of Statistics. This match finds deaths of South Australian cases who have died interstate, non-cancer deaths and any other deaths, which may have been missed otherwise. All Australian states have their data matched with the National Death Index.

#### **Hospital Ascertainment Checks Linkage**

Every six months, the SACR database is matched with the discharge data as requested from major Adelaide hospitals. This is done to detect cases of cancer known to the hospitals, but not known to the Cancer Registry. A locally written program is used to do this.

#### **Potential Duplicates Linkage**

Essentially, this is a link of the South Australian Cancer Registry database with itself to detect potentially duplicated records so that one copy can be deleted. "Automatch"<sup>6</sup> has been used recently to do this. A similar process is done at the national level for the accumulated data from all states on the National Cancer Statistics Clearing House database.

#### **BreastScreen SA Linkage**

BreastScreen SA is the South Australian population-based mammography breast screening program. A linkage of the BreastScreen SA database with the SACR database is done every six months to detect interval cancers. These are cancers which arise between the mammography screens done every two years. This is a quality assurance and accreditation measure for BreastScreen SA and is possible in South Australia because of the timeliness of the SA data. It is also occasionally the case that BreastScreen SA may have detected a case of cancer about which the Cancer Registry is unaware. Thus there is mutual benefit for both BreastScreen SA and the Cancer Registry. The matching is done with a locally written program using tight "Soundex" matching.

#### **Hospital-based (Clinical) Cancer Registries (HBCR's) Linkages**

South Australia has a number of HBCR's in most of the major hospitals. These registries are owned by the respective clinicians

and collect a wider range of data than does the population-based SACR. For example, data are collected with respect to cancer stage and grade, prognostic factors and types of first course of treatment given. With the express permission of the clinicians, HBCR data files intermittently are matched with the SACR for the mutual detection of cases of cancer missed by either party, or for informing the HBCR's of the existence of cases for them to chase in the hospitals. This is done using a locally written program.

#### **Other potential uses of data linkage in the SACR**

While the above indicates the routine uses of data linkage in the SACR, there are other linkages which might potentially be useful in enabling more extensive investigation of data available.

#### **Hospital-based (Clinical) Cancer Registries (HBCR's)**

If HBCR's were present in all hospitals, data linkage could enable the capture by the population-based cancer registry of clinical information, cancer grade and stage, investigations performed, prognostic factors and treatment modalities given. Australian population-based registries generally do not hold this information, though registries in some other parts of the world do, for example, in USA.

Some of these data would enhance the breadth and depth of analyses performed; for example, cancer survival by stage. Rather than their using some summary measure, or performing the very time-consuming task of collecting the data and holding it itself, the population-based registry could link with the HBCR's at the analysis phase to gain access to the accurate and complete hospital-collected data.

However, linking would be dependent on the appropriate clearances and permissions being given by the clinicians who own the HBCR's and by the owners of the population-based registry.

#### **Hospital clinical and administrative systems**

Linking the cancer registry with hospital and clinical data systems would enable the wider hospital experience of the person with cancer to be explored. This could enable better clinical and administrative quality assurance procedures to be implemented and evaluated for improving the quality of patient care.

#### **Medicare and the Pharmaceutical Benefits Scheme databases**

The community experience of a person with cancer could be explored more readily by linking the cancer registry with these databases. Often there is a concentration on hospital cancer care. However, much of the care of cancer patients also occurs in the community. These links could enable the investigation of these patterns and associations with other illnesses. Potentially, better ways of caring for cancer patients in the community might be able to be devised.

### **Molecular biology and bioinformatics data repositories**

These links might enhance the molecular characterisation of tumours and their genetic associations. This ultimately might enable more precise treatment targeting of tumours.

### **Research databases**

Such linkages could enable a mutual enrichment of information for both the cancer registry and the respective research projects.

### **Geo-spatial databases**

It is a relatively common occurrence that cancer registry data are used to investigate the geo-spatial distribution of cancer. For example, this might be for a research investigation of aetiological and ecological relationships with cancer, or it might be for the applied investigation of a postulated "cancer cluster" in the community. Linking with a geo-spatial database and then overlaying relevant factors onto the map or picture of the earth (as in the Kehole "Earthviewer" system – <http://www.earthviewer.com/index.html>) potentially will provide more easily interpretable information of any potential associations between cancer distribution patterns and the environment. Being able to do time-series investigations of these geo-spatial associations would add even more value.

### **Inhibitions to data linkage of Cancer Registry data with other databases**

#### **Legislative environment**

Under the current legislative arrangements, the South Australian Cancer Registry cannot share its information with anyone not covered under Section 64d. Thus unless the legislation is changed to include the other organisation under the legislation's umbrella, or some sort of special dispensation is arranged, linking using identified data cannot occur.

#### **Privacy and confidentiality issues**

Linking the cancer registry with other datasets could, or could be seen to compromise the privacy and confidentiality rights of the people recorded on each dataset. There is also the possibility of breaching provisions of specific privacy legislation, such as the Australian National Privacy Principles.

#### **Moral and ethical issues**

It could be argued that linking datasets has moral and ethical implications which might question the appropriateness of doing so. Discussion of this often involves the usual conflict between utilitarianism (using data to the greatest effect for the greater good of the community) versus autonomy (the individual's right to maintain control of his/her own information and not have it shared or linked if they do not want that to happen).

#### **Community acceptance or non-acceptance**

In the end, whether widespread data linkage of cancer registry

data with other datasets will occur or not will depend on whether the community will accept this as a reasonable activity. In a democracy, this is the final arbiter. However, this does imply that the community is informed and able to and encouraged to participate in the debate and the decision-making processes.

### **Future role of a Cancer Registry in a "linked world"**

The cancer registry most likely will retain a role as an expert collector of quality epidemiological cancer data. This is because it could be said that everybody should not try to do everything themselves. That is, the cancer registry should collect the minimum amount of data it requires to do its job. It should not try and collect every piece of clinical, psychological or social data about the person, when such data are collected elsewhere by people more expert in those areas.

However, to take better advantage of other data sources, the cancer registry would benefit from greater integration within a broad-based Health Information System. This is becoming more feasible technically with the exponential growth in the availability and functionality of information technology and the corresponding reduction in cost. Broadband Internet and emerging communications technologies (for example, wireless) will enhance these possibilities further. Using these technologies to link with other data sources has the potential to enable greater and more beneficial use of the cancer epidemiological data collected in the cancer registry.

If the community would accept a personal health identification number, linking of disparate datasets would become easier within the health system. However, it would not assist the linking of health data with datasets in other sectors, such as environmental systems or welfare systems.

### **Conclusions**

Cancer is a common, important and serious health problem in our community. Cancer registries provide essential population-based epidemiological surveillance information, which enables better cancer prevention, control and treatment activities to be put in place and evaluated. Data linkage is integral to cancer registry work in quality assurance activities, finding deaths, maximising ascertainment of cases and assisting others in their work. More widespread data linkage with other data sources and increased integration within broad health information systems could enhance the breadth and depth of information derived from the cancer registry. However, legislative settings, privacy and confidentiality concerns, moral and ethical issues and community acceptance are the most important limiting factors which will temper the more widespread linking of cancer registry data with others.

### **Postscript**

For a recent, comprehensive overview of record matching and data linkage methodologies, the reader is referred to the recent publication by Leicester Gill (2001)7.

## Acknowledgements

The South Australian Cancer Registry is located within and is fully funded by the South Australian Department of Human Services Statewide Division's Epidemiology Branch. This paper acknowledges and thanks the people who do the excellent day to day work of the registries: eg, the Registry staff: Maria Cirillo, Mary Merdo, Teresa Molik, Elaine Morton, Chris Scott, Heather Hall and Joanne Bell; Lesley Milliken (Coordinator of Hospital-based Registries); Dr Colin Luke (Clinical Epidemiology); Health Statistics Unit – Kevin Priest (who does all the data matching), Graeme Tucker and Anh- Minh Thi Nguyen; Information Management Services Division current and former staff who designed and implemented and/or who now maintain the Cancer Registries computer systems – such as Joanne McQuin, Stuart Speight, Alan Boman, Geoff Murphy, Andrew Grandison, Jayson Rogers, Som Liensavanh, Phil Ancell, and others. The former Epidemiology Branch and Cancer Registry Director A/Prof David Roder and A/Prof Kerry Kirke (Executive Director of the Anti-Cancer Foundation of South Australia) also are acknowledged for their past and/or continuing contribution to Cancer Registry activities.

## References

- South Australian Cancer Registry 2001. Epidemiology of cancer in South Australia. Incidence, mortality and survival 1977 to 2000. Incidence and mortality 2000. Analysed by type and geographical location. Twenty-four years of data. Adelaide: Openbook Publishers. <http://www.dhs.sa.gov.au/pehs/disease-control-stats.htm>
- Australian Institute of Health and Welfare (AIHW) & Australasian Association of Cancer Registries (AACR) 2001. Cancer in Australia 1998. AIHW cat. No. CAN 12. Canberra: AIHW. <http://www.aihw.gov.au/publications/can/ca98/index.html>
- Mathers C., Vos T., Stevenson C. 1999. The burden of disease and injury in Australia. AIHW cat. no. PHE 17. Canberra: AIHW.
- Jensen O.M., Parkin D.M., MacLennan R, Muir C.S. and Skeet R.G. (eds) 1991. Cancer Registration: Principles and Methods. IARC Scientific Publications No. 95. Lyon: International Agency for Research on Cancer World Health Organisation and International Association of Cancer Registries.
- Esteban D., Whelan S., Laudico A. and Parkin D.M. (eds) 1995. Manual for Cancer Registry Personnel. IARC Technical Report No. 10. Lyon: International Agency for Research on Cancer, World Health Organisation and International Association of Cancer Registries.
- MatchWare Technologies, Inc 1998. Automatch Generalized Record Linkage System. Maine: MatchWare Technologies, Inc.
- Gill L 2001. Methods for Automatic Record Matching and Linking and their use in National Statistics. National Statistics Methodology Series No. 25. London: Her Majesty's Stationery Office (HMSO).