

## 6 Methods and terminology

This section sets out the issues and terms used in cancer registration and highlights practices adopted by the SACR over its history. While this is not an extensive listing of all the issues it highlights those that are of significance.

Explanation of technical terms used in the report will assist the reader in interpreting the tables and graphs.

### 6.1 Coding issues in cancer registration

Demographic and cancer-related information is entered by Cancer Registry staff from the sources described and following the procedures explained in Section 1.2. The SACR currently use the International Classification of Diseases (9<sup>th</sup> Revision) to describe the topographic cancer site. The Systematized Nomenclature of Medicine and Modifications (SNOMED II) is used to classify histopathology. For particular cancer sites there are coding rules and issues which need to be highlighted:

#### Exclusions

The common basal and squamous cell skin cancers are not included in the cancer registry collections, with the exception of such lesions of skin immediately adjacent to the lip and anus.

“In-situ” cancers and neoplasms of uncertain behaviour are not included, except when otherwise stated in Section 1.2 of this report.

Sites for metastatic disease are not coded as such, but are assigned to the original primary sites or as “unknown primary” where applicable.

#### Coding options

- Soft-tissue cancers such as sarcomas and Merkel cell tumours of the intra-thoracic, breast or abdominal organs are coded to the relevant organ. Others are coded to the ICD-9 soft tissue site 171.
- Bladder tumours diagnosed prior to 1982 were all recorded as invasive cancers. These would have included some “in-situ” carcinomas and papillary non-invasive tumours.
- Urinary-tract tumours are counted as one primary when they are multifocal transitional cell carcinomas.
- Pre-leukaemic states such as myelodysplasia are recorded only when there is leukaemic change.
- Certain changes have been made to incorporate acute monocytic leukaemia and other specified leukaemias in the acute myeloid FAB classification.

## **6.2 Cancer registry terminology**

A number of terms are used to define concepts in cancer registry work. Some of the most common ones are discussed below.

### **Cancer incidence**

Cancer incidence is defined as the number of new cases of cancer notified for a specified period and for a specified population (e.g. South Australia for 2002 or 1977-2002). It is usually presented as either the number of new cases or as a rate (see below).

### **Cancer mortality**

Cancer mortality is defined as the number of deaths where cancer is specified as the underlying cause of death. The underlying cause of death is derived from the death certificate issued by a certified medical practitioner and uses the World Health Organization's rules for attribution of cause of death. Information about death and its cause may form part of the mandatory notification when cancer cases die in hospitals or is retrieved by linking the SACR to the Registry of Births, Deaths and Marriages. Non-cancer deaths are also recorded on the Registry however these are not coded to specifically to their allocated cause of death, only a non cancer death code.

### **Age-specific rates**

Age-specific incidence or mortality rates are calculated by dividing the number of new cases or deaths in each age group by the at risk population for that age group and multiplying by 100,000. These rates are usually presented in 5 year age groups and by sex.

### **Crude rates**

A crude incidence or mortality rate is defined as the number of new cancer cases or deaths (usually across all ages) divided by the population at risk in a specified time period. Crude incidence and mortality rates in this report were calculated using the estimated resident South Australian population for 2002 and are expressed as cases or deaths per 100,000 population per annum.

### **Age-standardised rates**

Summary incidence or mortality rates across all ages can be calculated to provide an overview of the impact of cancer. These rates are either expressed as crude rates (see above) or standardised rates (sometimes referred to as adjusted rates). These standardised rates enable comparisons of cancer rates between populations with different age distributions – an important factor as the risk of cancer increases with age.

Age-standardised incidence or mortality rates are measures of differences in cancer rates between populations, had their age distributions been the same.

The age-standardised method used in this report is the direct method where a Standard (or reference) population is used. This report presents rates standardised to the Australian 2001 Population and WHO World Standard Population (2001).

## Risk

Cumulative risk is the risk an individual would have of developing or dying of a particular cancer over a defined life span if that person were not to die beforehand from another cause. Cumulative risk is usually calculated using the following formula:

$$\text{Cumulative risk} = (1 - e^{-\text{cumulative rate}/100})$$

The cumulative rate (a component of the calculation) is the sum of the age-specific incidence or mortality rates over a certain specified age range (life expectancy or other specified range). It is calculated by using the formula:

$$\text{Cumulative rate} = \frac{5 \times (\text{sum of age specific rates}) \times 100}{100,000}$$

This formula assumes that age groups are arranged in 5 yearly blocks. Typically this rate ranges from less than 1 per cent for rare cancers to around nine per cent for common cancers.

Historically most cancer registries report this risk to age 74 (approximating the life expectancy) however life expectancy has increased markedly over the last decades and therefore this report also presents some measures of risk to later ages (see section 2.4).

Lifetime risk is another common way of expressing risk, where the risk is expressed as a 1 in x chance of being diagnosed with or dying of a cancer. It is calculated by the following formula:

$$\text{Lifetime risk} = 1 / \text{cumulative risk}$$

For example a cumulative rate of, say, 4.4 percent for lung cancer in males would mean that one out of every 23 males would be expected to be diagnosed with lung cancer by age 75 years if he were not to die before that age from another disease.

## Person-Years of Life Lost

Person-Years of Life Lost (PYLL) is a measure of the number of years of life lost per annum due to premature death from a particular cause given a specified life expectancy. As discussed above life expectancy has changed however this report adopts the international approach of reporting this measure to age 74. While there are a number of methods used to estimate this cancer impact this report uses the simple approximation of:

$$\text{PYLL} = 75 - \text{age at death} \times \text{number of deaths in at each age}$$

The calculation in this report is performed on 5 year age groups from 0-4 through to age 70-74.

### **6.3 Projections of cancer rates for selected sites**

The full time series of crude incidence rates were graphed separately for males and females for each site. From these graphs a subjective decision was made about which part of the series best reflected recent trends in incidence for these sites. For consistency, where the sites were in the top 15 for both males and females, the same part of the series was used for projection for each sex.

There was a discontinuity in the series for Prostate Cancer, which was reflected in the series for all sites for males. Thus cancers for all sites for males were only modelled based on data from 1996. The series for all cancers for females had no such problem, but for consistency this series was also modelled on data limited to 1996 onwards.

Key sites and appropriate years of diagnosis were selected as the base for projections. For each site, the age and sex specific crude rates were assessed for a linear trend. This was done by fitting a simple linear regression of the series against the year, and examining whether the coefficient of year was statistically significant at the 5% level. This model utilised the calculation of robust estimators of standard errors provided as an option by the statistical package “Stata”. The Huber/White/sandwich robust variance estimator produces consistent standard errors for Ordinary Least Squares regression coefficient estimates in the presence of heteroskedasticity.

Since the data are time series, it is possible that they are autocorrelated. Accordingly the errors in the regression model were tested for autocorrelation using the Durbin Watson alternative test for first order serial correlation in the disturbance. Where autocorrelation in the errors was detected, the model was refit using the Newey-West variance estimator, which handles autocorrelation up to and including a specified lag, as well as the presence of heteroskedasticity. The Newey-West model produces variance estimates that are exactly the Huber/White/sandwich robust variance estimates calculated by the robust regression estimator above when no lag is included in the model. Thus the two models used to assess linear trend are entirely consistent. Only first order correlation was allowed for in the fitting of the models.

Where a linear trend was established, projections were based on this trend. Where the trend was not statistically significant, projections were based on the mean crude rate of the series over that part of the series being applied.