Appendix 1: Methods

Objective

This analysis provides estimates of relative survival after people are diagnosed with cancer. The analysis provides estimates by age at diagnosis, year of diagnosis and five-year period during which diagnosis took place.

The analysis covers all new cases of cancer diagnosed in Australian residents between 1982 and 1997. In determining survival, the cancer cases have been matched against deaths from all causes registered in Australia from 1982 to 1999.

Relative survival analysis

Relative survival and cause-specific survival are two methods used to estimate the probability of surviving a specific disease (Estève et al. 1994). Cause-specific survival is used when cause of death is known with certainty. However, cause of death is not always easy to determine. Further, for most cancer patients, the risk of dying from other causes is not negligible and should be adjusted for when analysing their survival experience (Ederer et al. 1961). In contrast to cause-specific survival, relative survival does not require knowledge of the cause of death. For this national cancer survival project, not all causes of death were known with certainty. However, the fact of death was known and therefore relative survival methods were appropriate. Another important reason for using relative survival is that many morbid conditions contribute to a specific cause. For example, suicide may be the cause of death but may have been committed because of depression resulting from a diagnosis of cancer. Relative survival allows for the contribution of cancer to other causes even when a specific underlying cause is known.

Relative survival is defined as the ratio of the observed survival rate for a given cohort of patients to the expected survival rate (Ederer et al. 1961). The expected survival rate is the rate that the patient group should have experienced based on the life table of the general population from which they were diagnosed (Estève et al. 1990). A relative survival of less than 100% implies that the cohort survived for less time than would be expected for the general population. A relative survival of 100% implies that survival in the cohort is no different from that in the general population. A relative survival of greater than 100% generally arises from clinical trials where one cohort is provided with a superior treatment and thus outlives its comparison cohort. It can also arise when there is poor follow-up of cases and not all observed deaths in the cohort are recorded. A relative survival proportion greater than 100% in the confidence intervals generally happens when there are small numbers in the population being analysed, generating larger variances and increasing the size of the confidence intervals.

For example, in the general population during 1992–1997, the expected proportion of males aged 60–69 years who survive for the next five years is 90.6%. The observed survival rate after five years for males diagnosed with lung cancer at age 60–69 is 10.8%. The five-year relative survival proportion for males diagnosed with lung cancer at age 60–69 is the ratio of these two percentages (10.8/90.6), that is 0.119, or 11.9% (Table 8.1).

The relative survival analysis in this report was undertaken using the SAS statistical software functions as developed by the Mayo Foundation in 1994 (Therneau 1994). This code was developed by Terry Therneau using SAS Version 8.1. Staff in the Queensland Cancer Registry further adapted and developed the code for local use (Baade 2000b). AIHW staff then further developed the code to handle national-level data. The resultant program calculates expected survival using the life table method and estimates relative survival using a Cox proportional hazards regression (Estève et al. 1990).

Results using this method will produce estimates which will be slightly different from those produced by other cancer registries. Other registries used different relative survival packages—New South Wales and Western Australian cancer registries used the RELSURV package as developed by Hédelin (1995), and the South Australia cancer registry used the SURV2 package as developed by Voutilainen et al. (1998). The results will also be slightly different to those produced by the AIHW in an earlier *report—Breast Cancer Survival in Australian Women 1982–1994* (AIHW 1998) which also used RELSURV to produce its estimates.

Confidentiality

Strict confidentiality and privacy provisions apply to the National Cancer Statistics Clearing House (NCSCH) and the National Death Index (NDI). Restrictions on the use and release of information are included in State and Territory legislation controlling the operation of the Registries of Births, Deaths and Marriages and the cancer registries. Within the Australian Institute of Health and Welfare the data are protected under the *Australian Institute of Health and Welfare Act 1987*. The *Privacy Act 1988*, the Australian Public Service Regulations and the *Commonwealth Crimes Act 1914* also control the release of information by AIHW staff. Further, the AIHW maintains a secure physical and computer environment.

Applications to access data in either the NCSCH or the NDI must have AIHW Ethics Committee approval and strict controls are applied to the information provided. Ethics approval was sought and obtained for the survival analysis reported here.

Data sources

National Cancer Statistics Clearing House (NCSCH)

Each year the NCSCH receives from the eight State and Territory cancer registries data on cancer diagnosed in residents of Australia. This started with cases first diagnosed in 1982. The data provided to the NCSCH enable record linkage to be performed and the analysis of cancer by site and histology.

Data used in the relative survival analysis were for the period 1 January 1982 to 31 December 1997.

National Death Index (NDI)

The NDI is a database maintained by the AIHW. It contains data on all deaths that have occurred in Australia since 1980. It is continually being updated and is current to mid-2001 (at the time this report was written). The data contained in the NDI come from State and Territory Registrars of Births, Deaths and Marriages.

As part of normal NCSCH operating practices, the NCSCH is regularly linked to the NDI. This linkage is undertaken to assist State and Territory cancer registries to identify deaths occurring interstate or that were not notified to the cancer register.

Note that although the NDI is current to mid-2001, follow-up cancer survival analysis finished at 31 December 1999. This cut-off date provided at least two years of follow-up for the persons diagnosed with cancer during 1997.

Life tables

Life tables by sex and single-year ages (0–99 years) were obtained from the Australian Bureau of Statistics for Australia, for each of the States and Territories and for regions (RRMA) for each year from 1982 to 1997. The method used to calculate the life tables is outlined by the Australian Government Actuary (1999) in *Australian Life Tables 1995–1997*. The total number of deaths that occurred in each year by individual age (0 to 99) and sex were then linked to the respective populations to determine hazard rates.

Hazard rates

Estimation of relative survival requires hazard rates by single-year ages for each year of follow-up. These hazard rates, λ_x, were calculated from life table information using the formula:

$$\lambda_x = -\ln(1 - q_x)$$

where q_x is the probability of dying between exact ages x and $x + 1$ and is calculated using the following standard approximation:

$$q_x = \frac{M_x}{(1 + M_x(1 - a_x))}$$

where M_x is the age-specific death rate of persons aged x

a_x is the assumed fraction of a year lived by those who die during the year.

The following assumptions were made for a_x:

- a_0 = 0.9 because deaths among the very young in Australia tend to be concentrated early in the first year of life
- $a_1 — a_{99}$ = 0.5 because those who die in the year will live, on average, half of a year during that year.

Confidence intervals

Where indicators include a comparison between time periods and age groups, rates are presented with a 95% confidence interval. This is because the observed value of a rate may vary due to chance even where there is no variation in the underlying value of the rate. The 95% confidence interval represents a range over which variation in the observed rate is consistent with this chance variation. These confidence intervals can be used as an