

# Appendix A: Data sources and issues

## National Cancer Statistics Clearing House (NCSCCH)

Numbers of new cases of cancer from 1982 to 2001 used in the projections in this report are sourced from the National Cancer Statistics Clearing House (NCSCCH), a national database maintained by the AIHW.

Cancer is a notifiable disease in all Australian states and territories. The data are collected by cancer registries and include clinical and demographic information about people with newly diagnosed cancer. This information is obtained from hospitals, pathologists, radiation oncologists, cancer treatment centres and nursing homes.

The state and territory cancer registries are members of the Australasian Association of Cancer Registries (AACR) and provide data annually to the AIHW, which is responsible for the national collection of cancer incidence statistics through the NCSCCH. National statistics are currently available for all years from 1982 to 2001 and are published annually, the most recent publication being *Cancer in Australia 2001* (AIHW & AACR 2004).

## Cancers not included in the projections

The NCSCCH does not include non-melanoma skin cancers (NMSC), as they are not required to be reported to the cancer registries. These are the most common cancers diagnosed in Australia, but only rarely result in mortality (AIHW & AACR 2004). The most common forms of NMSC are basal cell carcinoma (BCC) and squamous cell carcinoma (SCC). In 2001, there were an estimated 112,000 females and 144,000 males treated for BCC and 47,000 females and 71,000 males treated for SCC (NCCI 2003).

These cancers are usually diagnosed and treated outside hospitals by general practitioners and dermatologists and in skin cancer clinics (NCCI 2003), but they are not legally notifiable and are not routinely registered by all state and cancer registries. The ICD-10 code (WHO 1992) for NMSC is C44 and these cancers have been excluded from the projections, owing to the lack of complete data.

A number of chronic myeloproliferative and myelodysplastic syndromes with ICD-10 codes in the range D45–D47, formerly considered to be disorders of uncertain behaviour, have recently become recognised as cancers. The NCSCCH does not yet have complete coverage of these cancers, for reasons outlined in the following section, so they have also been excluded from the projections.

## Impact of coding and reporting changes

Cancers are registered and coded using the International Classification of Diseases for Oncology (ICD-O); either the second edition, ICD-O-2 (WHO 1990) or the third edition, ICD-O-3 (WHO 2000). In both editions of ICD-O, neoplasms (tumours) are coded by both morphology (tumour type and behaviour) and topography (site).

The AACR has agreed that all state and territory cancer registries will eventually code in ICD-O-3. By the time of reporting 2001 cancer data, only four states and territories

(Australian Capital Territory, New South Wales, Tasmania and Western Australia) had made the transition to ICD-O-3 and a complete recode of all historical cancers is still being completed for some states and territories.

In this report, cancers are tabulated and reported using the ICD-10 codes (WHO 1992) shown in Table 2 for females and in Table 4 for males. ICD-O morphology and topography codes are summarised using a single code in ICD-10.

In ICD-O-2, all malignant neoplasms (cancers) corresponded to an ICD-10 code in the range C00–C97, and benign and in situ neoplasms and other neoplasms of unknown or uncertain behaviour corresponded to an ICD-10 code in the range D00–D48.

In ICD-O-3, some neoplasms previously considered to be of uncertain behaviour are now coded as malignant (for a list, refer to Threlfall & Thompson 2004, p. 61). In ICD-10, these neoplasms correspond to a code in the range D45–D47, which also includes some non-malignant neoplasms. ICD-10 codes in the range D45–D47 include polycythaemia vera (D45), myelodysplastic syndromes (D46) and other neoplasms of uncertain or unknown behaviour of lymphoid, haematopoietic and related tissue (D47).

Only cancers (malignant neoplasms) are included in the projections, and because the coverage of cancers in the range D45–D47 is not yet complete in the NCSCH, these conditions have not been included in the projections.

The four states that reported 2001 cancer data in ICD-O-3 represent a total of 47.7% of the 2001 Australian population. These states reported a total of 218 cancers with ICD-10 codes in the range D45–D47 in 2001.

For the states and territories that already code in ICD-O-3, rules for multiple primaries (which apply to persons with more than one neoplasm present) have resulted in some cancers that were formerly reported as leukaemia in ICD-10 (codes C91–C95) being no longer reported, because of prior incidence of one of the above conditions in the range D45–D47.

A study of cases registered in 2002 by the WA Cancer Registry (Threlfall & Thompson 2004, p. 61) found that in addition to the 221 cases of leukaemia reported, there were 12 further cases not reported because of prior incidence of one of the above conditions. So out of a total of 233 that would formerly have been reported in the range C91–C95 in ICD-10, 12 (5%) will no longer be reported. The transition to ICD-O-3 may lead to similar declines in the number of cancers reported in the range C91–C95 in other states and territories.

Conversely, a number of neoplasms, particularly ovarian neoplasms, considered to be malignant in ICD-O-2 are now considered to be of uncertain behaviour in ICD-O-3 (for a list, refer to Appendix 6 of ICD-O-3, WHO 2000, p. 240). In addition, some previously recorded papillary transitional cell neoplasms of the bladder may be reclassified as non-malignant in the transition to ICD-O-3 (Thursfield et al. 2004), although such neoplasms may later progress to become malignant.

As a result, the number of cases of some cancers, particularly ovarian cancer and bladder cancer (included in cancers of the urinary tract) may be currently overstated in the NCSCH and hence the projections for these cancers will also be overstated. Work has not yet been completed to quantify the effect (if any) of these changes.

## ABS population estimates and projections

Data on estimated resident population (ERP) from 1982 to 2003 used in this report were sourced from the Australian Bureau of Statistics (ABS 2003b). The AIHW preferred series (series 8) of the ABS population projections (ABS 2003c) was used to provide estimates of ERP from 2004 to 2011.

The AIHW has examined the most recent ABS population projections and has determined that series 8 is the most appropriate to use. This series assumes a continuation of past trends of rapidly declining rates of mortality. The alternative ABS mortality scenario assumes a slow-down in the rate of decline in mortality; there is no evidence of this occurring. Different ABS scenarios for fertility and migration will have little impact on the older age groups in which cancer incidence predominates.

Full details of the ABS population estimates and projections used are shown in the internet-only tables that accompany this report on the AIHW website <[www.aihw.gov.au](http://www.aihw.gov.au)>.

The standard populations used to produce age-standardised rates (see Appendix B) are shown in Table A1.

**Table A1: Standard populations: Australia 2001 and WHO 2000**

Age group	Australian Standard Population 2001 (A2001)		WHO World Standard Population 2000 (W2000)	
	Number	% of total	Number	% of total
0-4	1,282,357	6.6	8.86	8.9
5-9	1,351,664	7.0	8.69	8.7
10-14	1,353,177	7.0	8.60	8.6
15-19	1,352,745	7.0	8.47	8.5
20-24	1,302,412	6.7	8.22	8.2
25-29	1,407,081	7.2	7.93	7.9
30-34	1,466,615	7.6	7.61	7.6
35-39	1,492,204	7.7	7.15	7.1
40-44	1,479,257	7.6	6.59	6.6
45-49	1,358,594	7.0	6.04	6.0
50-54	1,300,777	6.7	5.37	5.4
55-59	1,008,799	5.2	4.55	4.5
60-64	822,024	4.2	3.72	3.7
65-69	682,513	3.5	2.96	3.0
70-74	638,380	3.3	2.21	2.2
75-79	519,356	2.7	1.52	1.5
80-84	330,050	1.7	0.91	0.9
85+	265,235	1.4	0.63	0.6
<b>Total</b>	<b>19,413,240</b>	<b>100.0</b>	<b>100.03</b>	<b>100.0</b>

Source: A2001—ABS 2003a; W2000—Ahmad et al. 2000.

# Appendix B: Statistical methods

## Cancer incidence rates and summary measures

For examples of how to calculate the following rates, see Appendix B of *Cancer in Australia 2001* (AIHW & AACR 2004). All rates are expressed as annual rates per 100,000 population.

### Age-specific rates

An age-specific rate is defined as the number of events for a specified age group over a specified period (for example, a year) divided by the total population at risk of the event in that age group. The age-specific cancer incidence rates for females and males in this report were calculated by dividing the number of new cases of cancer in a year by the corresponding estimated resident population (ERP) in the same age group at 30 June of that year.

### Crude rates

A crude rate is a summary rate across all age groups that is defined as the number of events divided by the total population at risk of the event over a specified period (for example, a year). For example, the crude cancer incidence rate for Australian males in 2001 is defined as the total number of new cases of cancer occurring among males in 2001 across all age groups divided by the total Australian male ERP at 30 June 2001.

### Age-standardised rates (ASR)

Age-standardised rates are summary rates across all age groups that enable comparisons to be made between populations which have different age structures. This report uses direct standardisation, in which the age-specific rates are multiplied by a constant population, which effectively removes the influence of the age structure on the summary rate.

In this report, rates are age-standardised to two standards, the 2001 Australian total ERP (ABS 2003a) as advocated by the National Health Data Committee (NHDC) for use in Australian health statistics and the new world standard population advocated by the World Health Organization (WHO) for use in international comparisons (Ahmad et al. 2000). These two standards are shown in Table A1 in Appendix A.

The method used for all the calculations of age-standardised rates consists of three steps:

- Step 1* Calculate the age-specific rate (as above) for each age group.
- Step 2* Calculate the expected number of cases in each 5-year age group by multiplying the age-specific rates by the corresponding standard population and dividing by 100,000, giving the expected number of cases.
- Step 3* To give the age-standardised rate, sum the expected number of cases across all age groups. Divide this sum by the total of the standard population used in the calculation and multiply by 100,000.

## Projection of the age-standardised rates

### Methods for projecting cancer incidence rates

Previous AIHW projections of cancer incidence (AIHW & AACR 1998; AIHW & AACR: Jelfs et al. 1996) relied on projecting the age-specific rates using separate models for each age group. Although this works well for the most common cancers, there can be problems with rarer cancers where the trends in the age-specific rates may be highly volatile, especially at younger ages where there are very few cases for most cancers. Extrapolating conflicting trends for different ages can also lead to a projected age distribution that is completely different from the historical age distribution for that cancer.

More recent unpublished projections undertaken by the AIHW tried to resolve these problems using methods adapted by Dr Chris Stevenson from the age-period interaction methods used by Coory and Armstrong (1998) for the projection of cancer incidence for New South Wales Area and Rural Health Services. These in turn had been derived from international projection methodology using linear and non-linear Poisson regression models developed by Dyba and Hakulinen (Hakulinen & Dyba 1994). Dyba and Hakulinen have since made some refinements to their approach (Dyba et al. 1997; Dyba 2000).

Other projections have been done using more complex age-period-cohort models (Taylor & McNeil 1997) or even multiple models with a model averaging approach (NZMOH 2002). Recent projections by the South Australian Cancer Registry (2005) have tried to improve on the original AIHW approach by incorporating tests for first-order autocorrelation and using robust variance estimators. A forthcoming paper by Clements et al. (2005) fits generalized additive models with a two-dimensional spline for age and period, to project lung cancer incidence for a number of countries.

### Functional data analysis

For the current projections, a functional data analysis (FDA) approach (Ramsey & Silverman 1997, 2001) has been used. In this approach the age distribution of the cancer is treated as a function and it is this function that is modelled and projected, rather than each of the age-specific rates. This approach is less susceptible to problems with low numbers of cases in the younger age groups than separate models for each age group.

A number of the projection methods outlined above were tried, and they produced similar results for the more common cancers, where there are large numbers of historical cases. For this project, projections by age group for a range of cancers were required, including some rarer cancers with only a small number of historical cases. As with the age-period models, the FDA approach uses as data the age-specific rates, rather than summary measures such as the age-standardised rates, and as a result produces projections for each age group as well as prediction intervals (see below) for each projected age-specific rate.

The FDA approach was chosen as it produces projections for rare cancers such as female mesothelioma that are stable in the sense that they are not highly susceptible to minor changes in either the data or the model parameters.

The FDA approach was applied by adapting computer routines written for the statistical software package R <[www.r-project.org](http://www.r-project.org)>, by Professor Rob Hyndman of Monash University (Hyndman & Ullah 2005). Using these routines, the historical age distribution function of cancer incidence is firstly smoothed, using regression splines, which can be shown to produce similar results to the earlier age-period methods (Heuer 1997).

The smoothed age distribution functions are then modelled using robust weighted principal components analysis to find a set of optimal (orthonormal) basis functions. Time series forecasting methods are then used to project the coefficients associated with these basis functions. The projected coefficients are then combined with the basis functions to produce projected age distributions of cancer incidence. Using the FDA approach, the projected age distribution can change gradually over time, which is why this approach can be considered to generalise the related Lee-Carter approach (Lee 2000; Lee & Carter 1992).

Most of the modelling for these projections was done on the log scale, with the exception of cancers with a strongly increasing trend, where the model on the log scale appeared to unreasonably exacerbate the trend. The projections for melanoma and the skin and lip group of cancers are examples of where the model was fitted on a linear rather than a log scale. Note that when the modelling is done on a log scale and the results transformed back to the original scale, prediction intervals will not necessarily be symmetrical about the projected value (the lower 95% prediction bound will usually be closer to the projection than the upper 95% prediction bound).

## **Prediction intervals**

The computer routines produce prediction intervals for the projected age-specific rates, and these prediction intervals have been combined to produce prediction intervals for the crude and age-standardised rates for all ages.

Prediction intervals are confidence intervals for predictions and can be used to assess the relative expected accuracy of the different projections. For example, in the projection for all cancers for males, the expected number of cases for 2011 is 63,087 with a 95% prediction interval that ranges from 58,122 to 68,752. The width of this prediction interval is 10,630 cases which is 16% of the expected number of cases.

Compare this with the same calculations for female eye cancer where the expected number of cases for 2011 is 131 with a 95% prediction interval from 89 to 195. The width of this interval is 106 cases which is 81% of the expected number of cases. The projection for all cancers for males is based on a total of 735,308 historical cases from 1982 to 2001, whereas the projection for female eye cancer is based on a total of 1,647 cases from 1982 to 2001, so it is not surprising that the projection for all cancers for males is expected to be relatively more accurate.

The prediction intervals reflect two components of variability in the smoothed age-specific rates: (1) one that reflects the volatility in the age-specific rates and (2) one that reflects increased uncertainty the further in time we try to project. This second component we would expect to be small early in the projection period, but to increase later in the projection period. For some of the projections, this second component appears to be negligible compared with the first component, and the prediction interval appears to be much the same width through the entire projection period from 2002 to 2011.

Dr Mark Clements of NCEPH, who reviewed the methodology, has suggested further work on the prediction intervals, outside the scope of the current project. Initial simulations suggest that in most instances the current prediction intervals are conservative; in particular, they are too wide in the early part of the projection period, but appear suitable towards the end of the projection period.

95% prediction intervals have been calculated as this is the AIHW standard for confidence intervals. However, Professor Hyndman suggests the use of much narrower 80% intervals when dealing with predictions and this is set as the default in his computer package.

# References

- ABS (Australian Bureau of Statistics) 2003a. Australian demographic statistics. ABS cat. no. 3101.0, September quarter 2002, 2001 census edition – final. Canberra: ABS.
- ABS 2003b. Population by age and sex, Australian states and territories. ABS cat. no. 3201.0. Canberra: ABS.
- ABS 2003c. Population projections, Australia 2002 to 2101. ABS cat. no. 3222.0. Canberra: ABS.
- Ahmad OB, Boschi-Pinto C, Lopez AD, Murray CJ, Lozano R & Inoue M 2000. Age standardization of rates: a new WHO standard. EIP/GPE/EBD discussion paper series no. 31 <[www3.who.int/whosis/discussion\\_papers/discussion\\_papers.cfm](http://www3.who.int/whosis/discussion_papers/discussion_papers.cfm)>. Geneva: World Health Organization.
- AIHW (Australian Institute of Health and Welfare) 1998. Breast and cervical cancer screening in Australia 1996–1997. AIHW cat. no. CAN 3. Canberra: AIHW (Cancer Series no. 8).
- AIHW 2004. Cervical screening in Australia 2001–2002. AIHW cat. no. CAN 22. Canberra: AIHW (Cancer Series no. 27).
- AIHW 2005a. 2004 National Drug Strategy Household Survey: first results. AIHW cat. no. PHE 57. Canberra: AIHW (Drug Statistics Series no. 13).
- AIHW 2005b. BreastScreen Australia monitoring report 2001–2002. AIHW cat. no. CAN 24. Canberra: AIHW (Cancer Series no. 29).
- AIHW: Ridolfo B & Stevenson C 2001. The quantification of drug-caused mortality and morbidity in Australia, 1998. AIHW cat. no. PHE 29. Canberra: AIHW (Drug Statistics Series no. 7).
- AIHW & AACR (Australasian Association of Cancer Registries) 1998. Cancer in Australia 1991–1994 (with projections to 1999). AIHW cat. no. CAN 2. Canberra: AIHW (Cancer Series no. 7).
- AIHW & AACR 2003. Cancer in Australia 2000. AIHW cat. no. CAN 18. Canberra: AIHW (Cancer Series no. 23).
- AIHW & AACR 2004. Cancer in Australia 2001. AIHW cat. no. CAN 23. Canberra: AIHW (Cancer Series no. 28).
- AIHW & AACR: Jelfs P, Coates M, Giles G, Shugg D, Threlfall T, Roder D et al. 1996. Cancer in Australia 1989–1990 (with projections to 1995). Canberra: AIHW (Cancer Series no. 5).
- Booth H, Maindonald J & Smith L 2001. Age–time interactions in mortality projection: applying Lee–Carter to Australia. Working Papers in Demography 85 <[demography.anu.edu.au/Publications/workingpapers.php](http://demography.anu.edu.au/Publications/workingpapers.php)>. Canberra: Australian National University.
- Booth H, Maindonald J & Smith L 2002. Applying Lee–Carter under conditions of variable mortality decline. *Population Studies* 56:325–36.
- Clements MS, Armstrong BK & Moolgavkar SH 2005. Lung cancer rate predictions using generalized additive models. *Biostatistics* (in press).
- Coory M & Armstrong BK 1998. Cancer incidence projections for area and rural health services in New South Wales. Sydney: NSW Cancer Council.
- Coory M, Baade P, Aitken J, Smithers M, McLeod GRC & Ring I 2005. Trends for in-situ and invasive melanoma in Queensland, Australia, 1982 to 2002. *Cancer Causes Control* (in press).

- DoHA (Department of Health and Ageing) 2005. Investing in Australia's health: strengthening cancer care. Budget 2005–2006 Health Fact Sheet no. 1 <[www.health.gov.au/budget2005](http://www.health.gov.au/budget2005)>. Canberra: DoHA.
- Dyba T 2000. Precision of cancer incidence projections based on Poisson distributed observations. Helsinki: Finnish Statistical Society.
- Dyba T, Hakulinen T & Paivarinta L 1997. A simple non-linear model in incidence projection. *Statistics in Medicine* 16:2297–309.
- Erbas B, Hyndman RJ & Gertig DM 2005. Forecasting age-specific breast cancer mortality using functional data models. *Statistics in Medicine* (in press).
- Frazer I 2004. Preventing cervical cancer: the immunologist's role. Paper presented at the 18th World Conference on Health Promotion and Health Education. Melbourne, 26–30 April 2004.
- Giles GG, Armstrong BK, Burton RC, Staples MP & Thursfield VJ 1996. Has mortality from melanoma stopped rising in Australia? Analysis of trends between 1931 and 1994. *BMJ (British Medical Journal)* 312(7039):1121–5.
- Hakulinen T & Dyba T 1994. Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine* 13:1513–23.
- Heuer C 1997. Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* 53(1):161–77.
- Hill DJ, White VM & Gray NJ 1988. Measures of tobacco smoking in Australia 1974–1986 by means of a standard method. *Medical Journal of Australia* 149:10–2.
- Hyndman RJ & Ullah MS 2005. Robust forecasting of mortality and fertility rates: a functional data approach. Invited paper presented at the 55th session of the International Statistical Institute <[www-personal.buseco.monash.edu.au/~hyndman](http://www-personal.buseco.monash.edu.au/~hyndman)>. Sydney, April 2005.
- Lee RD 2000. The Lee–Carter method for forecasting mortality, with various extensions and applications. *North American Actuarial Journal* 4(1):80–91.
- Lee RD & Carter LR 1992. Modeling and forecasting US mortality. *Journal of the American Statistical Association* 87(419):659–71.
- Lee R & Miller T 2001. Evaluating the performance of the Lee–Carter method for forecasting mortality. *Demography* 38(4):537–49.
- Leigh J & Driscoll T 2003. Malignant mesothelioma in Australia, 1945–2002. *International Journal of Occupational and Environmental Health* 9:206–17.
- NCCI (National Cancer Control Initiative) 2003. The 2002 national non-melanoma skin cancer survey: a report by the NCCI Non-melanoma Skin Cancer Working Group. Ed. Staples MP. Melbourne: NCCI.
- NOHSC (National Occupational Health and Safety Commission) 2004. The incidence of mesothelioma in Australia 1999 to 2001, Australian Mesothelioma Register report 2004. <[www.nohsc.gov.au/PDF/Statistics/Meso2004.pdf](http://www.nohsc.gov.au/PDF/Statistics/Meso2004.pdf)>. Canberra: NOHSC.
- NZMOH (New Zealand Ministry of Health) 2002. Cancer in New Zealand – trends and projections. *Public Health Intelligence Occasional Bulletin* no. 15. Wellington: NZMOH.
- Ramsey JO & Silverman BW 1997. *Functional data analysis*. New York: Springer-Verlag.
- Ramsey JO & Silverman BW 2001. *Functional data analysis*. In: Smelser J & Bates PB (eds). *International encyclopedia of the social & behavioural sciences*. Oxford: Pergamon. <[www.stats.ox.ac.uk/~silverma/pdf/iesbs.pdf](http://www.stats.ox.ac.uk/~silverma/pdf/iesbs.pdf)>:5822–8.
- SACR (South Australian Cancer Registry) 2005. Cancer in South Australia 2002 – with projections to 2005. Adelaide: South Australian Department of Health.

Taylor R & McNeil D 1997. Predictions of incidence of major cancers in NSW to 2001. Sydney: NSW Cancer Council.

Threlfall TJ 1997. Cancer incidence and mortality projections for Western Australia, 1996–2001. Statistical Series no. 50. Perth: Department of Health, Western Australia.

Threlfall TJ & Thompson JR 2004. Cancer incidence and mortality in Western Australia, 2002. Statistical Series no. 71. Perth: Department of Health, Western Australia.

Thursfield V, Farrugia H, Billett D & Giles G 2004. Cancer in Victoria 2003. Canstat no. 41. Melbourne: Cancer Council Victoria.

Tracey E, Roder D, Bishop J, Chen S & Chen W 2005. Cancer in New South Wales: incidence and mortality 2003. Sydney: Cancer Institute NSW.

Tracey EA & Supramaniam R 2002. Cancer in New South Wales: incidence and mortality 2000 featuring projections to 2010 for selected sites. Sydney: Cancer Council NSW.

White V, Hill D, Siahpush M & Bobevski I 2003. How has the prevalence of cigarette smoking changed among Australian adults? Trends in smoking prevalence between 1980 and 2001. *Tobacco Control* 12(supplement ii):67–74.

WHO (World Health Organization) 1990. International classification of diseases for oncology, 2nd edn. Eds: Percy C, Van Holten V & Muir C. Geneva: WHO.

WHO 1992. International statistical classification of diseases and related health problems, 10th revision. <[www3.who.int/icd/vol1htm2003/fr-icd.htm](http://www3.who.int/icd/vol1htm2003/fr-icd.htm)>. Geneva: WHO.

WHO 2000. International classification of diseases for oncology, 3rd edn. Eds: Fritz A, Percy C, Jack A, Shanmugaratnum K, Sobin L, Parkin DM & Whelan S. Geneva: WHO.

Woodward SD 1981. Trends in cigarette consumption in Australia. *Australia and New Zealand Journal of Medicine* 14:405–7.