

# Appendix A: Data and statistical issues

## Data sources

Multiple data sources were analysed to produce this report. These are summarised in Table A1. All data used report are based on calendar years.

**TableA1: Sources for data presented in this report**

Indicator	Description	Data source
1	Participation	BreastScreen Australia state and territory services
2	Cancer detection	BreastScreen Australia state and territory services
3	Sensitivity	BreastScreen Australia state and territory services
4	DCIS detection	BreastScreen Australia state and territory services
5	Recall to assessment	BreastScreen Australia state and territory services
6	Rescreening	BreastScreen Australia state and territory services
7	Incidence (ICD-9 174)	National Cancer Statistics Clearing House, AIHW
8	Mortality (ICD-9 174, ICD-10 C50)	National Mortality Database, AIHW

## Population data

The Australian Bureau of Statistics' estimated resident population (ERP) data were used to calculate screening participation and cancer incidence and mortality rates.

Participation rates were calculated using the average of the 1999 and 2000 estimated resident female populations. The only exceptions to this were participation rates by socioeconomic status, by language spoken at home and by Indigenous status. The population data on which the participation rates are based are found on the web site of the Australian Institute of Health and Welfare (<http://www.aihw.gov.au>).

The ABS does not calculate ERP by socioeconomic status or language spoken at home. In these cases the denominator was calculated by applying the age-specific distribution of socioeconomic status and language spoken at home from the 1996 ABS Census to the relevant age specific ERP counts. The most recent direct count of the Aboriginal and Torres Strait Islander population was carried out in the 1996 Census. However, the ABS has released estimates of Aboriginal and Torres Strait Islander population for more recent years. These estimates were used as the denominator for rates for Indigenous people.

## Statistical analysis of BreastScreen monitoring indicators

### Crude rates

A crude rate is defined as the number of events over a specified period (for example, a year) divided by the total population at risk of the event. For example, a crude cancer incidence rate is defined as the number of new cases of cancer in a specified period of time divided by the population at risk.

### 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. Age-specific rates in this report were calculated by dividing the number of deaths, cancer cases or women participating in the screening programs in each specified age group by the corresponding population in the same age group.

### Age-standardised rates (ASR)

Age-standardised rates enable comparisons to be made between populations with different age structures. This publication used direct standardisation, in which the age-specific rates are multiplied by a constant population. This effectively removes the influence of the age structure on the summary rate.

The National Health Data Committee has advocated the use of the 1991 Australian total estimated resident population as the standard population until the year 2001. Mortality, incidence and participation rates are age-standardised to the 1991 Australian total population. For statistics based on the population of women screened – that is, cancer detection rates, interval cancer rates and program sensitivity – the rates are standardised to the 1998 population of women screened by BreastScreen Australia. The standard populations used in this report are found on the AIHW web site (<http://www.aihw.gov.au>).

The method used for all of these calculations consists of three steps:

Step 1: Calculate the age-specific rate 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 the appropriate factor (i.e. 100,000 for mortality and incidence rates, 10,000 for cancer detection and sensitivity rates and 100 for the participation rate).

Step 3: Sum the expected number of cases in each age group, divide by the total of the standard population and multiply by the appropriate factor (i.e. 100,000 for mortality and incidence rates, 10,000 for cancer detection and sensitivity rates and 100 for the participation rate). This gives the age-standardised rate.

### Rate denominators

Death rates and cancer incidence rates are expressed in this report as annual rates per 100,000 population. Rates for cancer detection are calculated per 10,000 women screened. Screening participation rates are expressed as a percentage of the eligible population. Rescreen and recall to assessment rates are expressed as a percentage of women screened.

## **Confidence intervals**

The 95% confidence intervals in this report were calculated using the software package Palisade @Risk (<http://www.palisade.com>). The calculations were based on 1000 simulations using a binomial or Poisson distribution with parameters calculated from the observed data. The confidence intervals represent a range of values within which the true value of the rate is likely to fall in 95% of iterations.

The confidence intervals are used to provide an approximate indication of the differences between rates. Where the confidence intervals of two rates do not overlap, the corresponding rates are statistically significantly different from each other. This is used to compare individual stratified rates with the all-Australia rate. To be truly rigorous, such a comparison should be between a given rate and the rate calculated from the all-Australia data excluding the data underlying the specific rate in the comparison. Presentation of such a comparison in this report would, however, be unnecessarily complex. The approximate comparisons presented might understate the statistical significance of some differences, but they are sufficiently accurate for the purposes of this report.

As with all statistical comparisons, care should be exercised in interpreting the results of the comparison. If two rates are statistically significantly different from each other, this means that the difference is unlikely to have arisen by chance. Judgement should, however, be exercised in deciding whether or not the difference is of any practical significance.

## **BreastScreen Australia Data Dictionary**

A data dictionary has been developed for the BreastScreen Australia Program (AIHW & DoHA forthcoming). Summary definitions of key concepts and terminology used in this report are given in the glossary. More detailed definitions and explanations may be found in the data dictionary.

## **Stratification variables**

The data in this report are presented stratified by the age of the women at the time of screening (for the screening data), at the time of diagnosis (for the cancer incidence data) or at the time of death (for the cancer mortality data). A number of other stratification variables apply to some or all of the data presented:

- state or territory;
- geographic location;
- socioeconomic status;
- Indigenous status;
- main language spoken at home;
- tumour size; and
- screening round.

## **State or territory**

The state or territory reported is the one where screening took place (for the screening data) or where the diagnosis was made (for the cancer incidence data) or where the death was registered (for the cancer mortality data).

This means that it is possible for a woman to be double-counted in the screening data. If she was screened in one jurisdiction and then screened again less than two years later in another jurisdiction, both screens may be included in the participation rate. This should, however, have a negligible effect on the reported participation rates.

## Geographic location

Analysis of participation and mortality data by geographic area uses the Rural, Remote and Metropolitan Areas (RRMA) classification. This classification, developed in 1994 by the then Department of Primary Industries and Energy and the then Department of Human Services and Health, is used as a framework for examining breast cancer screening participation and mortality data at the national level. Seven categories are included – two metropolitan, three rural and two remote zones (see Table A2).

**Table A2: Structure of the Rural, Remote and Metropolitan Areas classification**

Zone	Category
Metropolitan	Capital cities
	Other metropolitan centres (urban centre population > 100,000)
Rural	Large rural centres (urban centre population 25,000–99,999)
	Small rural centres (urban centre population 10,000–24,999)
	Other rural areas (urban centre population <10,000)
Remote	Remote centres (urban centre population > 5,000)
	Other remote areas (urban centre population <5,000)

A more recent geographic classification system has been developed using 1996 Population Census data. This system, known as the Accessibility/Remoteness Index of Australia (ARIA) categorises areas according to their distance from 'service centres'. Service centres are urban centres with a population of 5,000 or more as at the 1996 Census. The ARIA system classifies areas as highly accessible, accessible, moderately accessible, remote and very remote.

The ARIA system was not designed as a replacement for the RRMA classification. In particular, it does not allow a comparable categorisation. Accessibility is judged purely on distance to an urban centre. For example, Albury, Tamworth and Inner Sydney all have the same classification (highly accessible) using the ARIA system. An updated rural/remote/metropolitan categorisation is being developed using ARIA. Until that new categorisation becomes available, the existing RRMA system will continue to be used.

The RRMA classification is based on statistical local areas (SLA) and allocates each SLA in Australia to a category based primarily on population numbers and an index of remoteness (DPIE & DSHS 1994). Both the size of SLAs and the distribution of population within them vary considerably. This can mean that within a remote SLA there are pockets that are rural rather than remote and vice versa.

The use of SLAs for coding geographic regions is not straightforward. In particular, SLA boundaries change over time. Coding data to SLAs thus raises difficulties with tracking these changes over time and ensuring that all data are coded consistently. Instead, for this report the data were coded to postcode and a concordance was developed to map postcode to

RRMA. This mapping is not exact, since SLA boundaries may cross postcode boundaries. As a result, the proportion of each postcode that could be mapped to each RRMA category was calculated and used to code the data.

A total of 262 of the 1999–2000 and 507 of the 1998–1999 BreastScreen screening data records had postcodes that could not be mapped to an RRMA classification. These were allocated proportionally between the classifications.

### **Socioeconomic status**

Socioeconomic status was coded according to the Index of Relative Socio-economic Disadvantage (IRSD). The IRSD is one of the socioeconomic indexes for areas (SEIFA indexes) developed by the Australian Bureau of Statistics to categorise geographic areas according to their social and economic characteristics.

It is important to note that the IRSD relates to the average disadvantage of all people living in a geographic area. Hence any variability between groups based on the IRSD will probably be smaller than if the variability had been measured between individuals.

Like the RRMA classification, the IRSD was developed for SLAs. However, as with the RRMA coding, for this report the data were coded to postcode and a concordance was developed to map postcode to quintile of IRSD. Again, this mapping is not exact, since SLA boundaries can cross postcode boundaries. As a result, the proportion of each postcode that could be mapped to each IRSD quintile was calculated and used to code the data.

A total of 294 of the 1999–2000 and 354 of the 1998–1999 BreastScreen screening data records had postcodes that could not be mapped to an IRSD quintile. These were allocated proportionally between the quintiles.

### **Indigenous status**

The BreastScreen Australia Data Dictionary (AIHW & DoHA forthcoming) specifies that Indigenous status should be coded as

- Aboriginal;
- Torres Strait Islander;
- both Aboriginal and Torres Strait Islander;
- not Indigenous; or
- not stated.

For the purposes of this report these categories were amalgamated and the data stratified into three categories:

- Indigenous;
- not Indigenous; or
- not stated.

Some jurisdictions do not use the 'Not stated' category. These jurisdictions classify women who do not record their Indigenous status as 'Not Indigenous' rather than 'Not stated'. This means that analyses based upon Indigenous status should be interpreted with caution.

## **Main language spoken at home**

The BreastScreen Australia Data Dictionary (AIHW & DoHA forthcoming) recommends that main language spoken at home be coded according to the four-digit ABS Australian Standard Classification of Languages, 1998 (ABS cat. no. 1267.0). This report has collapsed the classification into the simple dichotomy of 'English' and 'Other language'. If main language spoken at home is not given, this is coded as 'Not stated'.

Although this stratification is reported as 'main language spoken at home', practice varies between the jurisdictions as to how this information is collected. In some jurisdictions there may thus be some lack of comparability with the BreastScreen Data Dictionary definition of 'main language'.

In addition, some jurisdictions do not use the 'Not stated' classification. If main language spoken at home is not given, it is set to a default value. The default used is not the same for all jurisdictions. This means that the analysis based upon main language spoken at home should be interpreted with caution.

## **Tumour size**

Tumour size is the size in millimetres of the malignant lesion, and applies to invasive cancers only. For more details about this stratification, see the definition given in the BreastScreen Australia Data Dictionary (AIHW & DoHA forthcoming).

## **Screening round**

The BreastScreen Australia Data Dictionary distinguishes between a woman's screening round in the national program and her round in the state or territory program. Round in the national program is used for this stratification in this report. However, it is not always possible to determine round in the national program, so for some women this stratification has been collected as round number in the state or territory program.

## **Mortality data**

Mortality data in this report are given for 1987–2000. During this time, changes have been made to the coding and processing of mortality data that affect the comparability of the data. Data holdings for 1987–1996 were coded using the ninth revision of the International Classification of Diseases (ICD-9). These data were coded manually. Data holdings for 1997 onwards have been coded using ICD-10. These data were coded using an automated system with slightly different coding rules.

The change to the coding and processing of these data has introduced a break in the mortality data time series. The Australian Bureau of Statistics developed comparability factors, which are applied to the pre-1997 data, so that a single time series may still be derived (ABS 2002). For breast cancer, the comparability factor is close to one.

The application of a comparability factor causes the number of deaths prior to 1997 to be non-integer. Rounding has been used to put the number of deaths into whole numbers.



# Abbreviations

**AACR:** Australasian Association of Cancer Registries

**ABS:** Australian Bureau of Statistics

**ACT:** Australian Capital Territory

**AIHW:** Australian Institute of Health and Welfare

**AHMAC:** Australian Health Ministers Advisory Council

**ASR:** age-standardised rate

**ASR(A):** age-standardised rate – standardised to the Australian standard population

**BSA:** BreastScreen Australia

**BSANAC:** BreastScreen Australia National Advisory Committee

**CI:** confidence interval (see glossary)

**DoHA:** Australian Government Department of Health and Ageing

**DCIS:** ductal carcinoma in situ

**DHSH:** Department of Human Services and Health (1994 to 1996)

**ERP:** estimated resident population

**NBCC:** National Breast Cancer Centre

**NHS:** National Health Survey

**NQMC:** National Quality Management Committee

**NSW:** New South Wales

**NT:** Northern Territory

**Qld:** Queensland

**SA:** South Australia

**SES:** socioeconomic status

**SLA:** statistical local area

**Tas:** Tasmania

**Vic:** Victoria

**WA:** Western Australia

**WHO:** World Health Organization

# Glossary

**Administrative databases:** observations about events that are routinely recorded or required by law to be recorded. Such events include births, deaths, hospital separations and cancer incidence. Administrative databases include the National Mortality Database, the National Hospital Morbidity Database and the National Cancer Statistics Clearing House Database.

**Age-specific rate:** a rate for a specific age group. The numerator and denominator relate to the same age group.

**Age-standardised rate:** weighted average of age-specific rates according to a standard distribution of the population by age to eliminate the effect of different age distributions and thus facilitate valid comparison of groups with differing age compositions.

**Assessment:** further investigation of a mammographic abnormality or symptom reported at screening. This includes women who choose assessment outside the Program.

**Benign:** not cancerous.

**Cancer (malignant neoplasm):** a term used to describe one of several diseases that result when the process of cell division, by which tissues normally grow and renew themselves, becomes uncontrolled and leads to the development of malignant cells. These cancer cells multiply in an uncoordinated way, independently of normal growth control mechanisms, to form a tumour. The tumour can expand locally by invasion or systemically by metastasis via the lymphatic or vascular systems. If left untreated, most malignant tumours eventually result in death.

**Cancer death:** a death where the underlying cause is indicated as cancer. People with cancer but dying of other causes are not counted in the death statistics in this publication.

**Confidence interval:** a range determined by variability in data, within which there is a specified (usually 95%) chance that the true value of a calculated parameter (for example, relative risk) lies.

**Core biopsy:** removal of a cylindrical sample of breast tissue under a local or general anaesthetic through a needle for microscopic examination.

**Data:** refers to the building blocks of health information, including observations from administrative databases and health survey data sets.

**Ductal carcinoma in situ:** a non-invasive tumour of the mammary gland (breast) arising from cells lining the ducts.

**Early review:** a woman is screened but not cleared for routine rescreening and instead is referred for further assessment within 6 to 12 months of the index screen.

**Epidemiology:** the quantitative study of the distribution and determinants of health-related states and events in populations and the application of this study to the control of health problems.

**False negative:** means that the test has incorrectly observed that the disease is not present.

**False positive:** means that the test has incorrectly observed that the disease is present.

**Film reading:** viewing of a radiographic depiction of the breast (a mammogram) to determine the presence or absence of an abnormality indicative of a tumour.

**Fine needle aspiration biopsy:** the sampling of cells from breast tissue for examination by a pathologist.

**Incidence:** see **New cancer case**.

**Index screening year:** the year for which the interval cancer rate and the program sensitivity rate are determined.

**Index screens:** all screening examinations performed within the index screening year.

**Indicators:** observations about data that have been analysed to provide a means of comparing measures of health within and between population groups.

**Indigenous:** a person of Aboriginal and/or Torres Strait Islander descent who identifies as an Aboriginal and/or Torres Strait Islander person and is accepted as such by the community with which he or she is associated.

**Information:** observations about data that have been analysed to provide a means of comparing measures of health within and between population groups.

**International Classification of Diseases:** WHO's internationally accepted classification of death and disease. The tenth revision (ICD-10) is currently in use.

**Interval cancer – invasive** (as defined for national reporting purposes by Kavanagh et al. (1999), with minor changes pending endorsement by the National Advisory Committee):

- an invasive breast cancer diagnosed after completion of a negative screening episode and before the next screening examination (within 24 months from the date of the previous screen).
- a case of invasive breast cancer that is diagnosed at early review or in the interval between assessment and early review, where the recommendation for early review is six months or more from the screening date.
- breast cancer diagnosed in a woman by BreastScreen Australia within 24 months of a negative screen (early rescreen) if the woman presents with a breast lump and/or clear or blood-stained nipple discharge in the breast in which the breast cancer was diagnosed; or
- an invasive breast cancer diagnosed between six and 24 months after a recommendation for assessment is made and a woman fails to attend assessment.

**Invasive cancer:** a tumour whose cells have invaded healthy or normal tissue.

**Lymph node:** masses of lymphatic tissue, often bean-shaped, that produce lymphocytes and through which lymph filters. These are located throughout the body.

**Mammogram:** a radiographic depiction of the breast.

**Metastasis:** the process by which a disease is transferred from one part of the body to another – for example, via the lymphatic system or the bloodstream.

**Mortality:** see **Cancer death**.

**New cancer case:** a person who has a new cancer diagnosed for the first time. One person can have more than one cancer and therefore may be counted twice in incidence statistics if it is decided that the two cancers are not of the same origin. This decision is based on a series of principles set out in more detail in a publication by Jensen et al. (1991).

**Next scheduled screening examination:** 24 months after previous screen unless the woman is recommended for annual rescreening, when the next scheduled screening examination is 12 months.

**Population estimates:** official population numbers compiled by the Australian Bureau of Statistics at both state and territory and statistical local area levels by age and sex, as at 30

June each year. These estimates allow comparisons to be made between geographic areas of differing population sizes and age structures.

**Prevalence:** the number of instances of a specific disease or other condition in a given population at a designated time.

**Recruitment:** strategies that aim to promote participation of women in the BreastScreen Australia Program through direct contact with women in the target age group and education of health practitioners and the general public. Women are encouraged to attend every two years.

**Rescreening:** the next screening examination after the screening episode in the index screening year.

**Risk factor:** an attribute or exposure that is associated with an increased probability of a specified outcome, such as the occurrence of a disease. Risk factors are not necessarily the causes of disease.

**Screening:** the performance of tests on apparently well people in order to detect a medical condition at an earlier stage than would otherwise be the case. As a screening test is not intended to be diagnostic, so a person with a positive or suspicious result must be referred for diagnosis and treatment.

**Screening episode:** includes screening examination and assessment. Early review within 6–12 months of an initial screen is not considered part of the screening episode.

**Screening round:** the first screening round is a woman's first visit to a mammography screening service; a subsequent screening round means that she has been screened before. If she attends for the fourth screening round, she has been screened three times before.

**Screening round (first):** a woman's first visit to a BreastScreen Australia mammography screening service.

**Screening round (subsequent):** a woman's visit to a BreastScreen Australia mammography screening service when she has attended such a service before.

**Sensitivity:** the proportion of people with a disease who have a positive test result for the disease.

**Significant difference:** where rates are referred to as significantly different, or one rate is deemed significantly higher or lower than another, these differences are statistically significant. Rates are deemed statistically significantly different when their confidence intervals do not overlap, since their difference is greater than what could be explained by chance. See 'confidence intervals' in Appendix A for more information.

**Symptom:** any evidence of disease apparent to the patient. For the purposes of this report, symptoms refer to a self-reported breast lump and/or blood-stained or watery nipple discharge.

**Torres Strait Islander:** a person of Torres Strait Islander descent who identifies as a Torres Strait Islander and is accepted as such by the community in which he or she lives.

**Ultrasound:** diagnostic method based on the reflection of ultrasonic sound waves generated through scanning of, in this case, the breast. The reflections are viewed on a computer screen or photograph and checked for variations in images.

**Unit record file:** observations containing person-specific records from health surveys and administrative databases that are unanalysed and not tabulated. This is the most basic form of data and cannot be accessed for general use without appropriate confidentiality measures being in existence.

**Women-years at risk:** all women screened in the index screening year who are resident in the state or territory in which they are screened who have not reported a personal history of breast cancer.

# Bibliography

- ABS (Australian Bureau of Statistics) 2002. Causes of Death, Australia. 2000. Cat. no. 3303.3 ABS. Canberra: ABS.
- AHMAC (Australian Health Ministers' Advisory Council) Breast Cancer Screening Evaluation Committee 1990. Breast cancer screening in Australia: future directions. Australian Institute of Health: Prevention Program Evaluation Series No. 1. Canberra: Australian Government Publishing Service.
- 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 (Australian Institute of Health and Welfare) 2000. BreastScreen Australia Achievement Report 1997–1998. AIHW Cat. No. CAN 8. Canberra: AIHW (Cancer Series no. 13).
- AIHW (Australian Institute of Health and Welfare) & AACR (Australasian Association of Cancer Registries) 2002. Cancer in Australia 1999. AIHW cat. no. CAN 15. Canberra: AIHW (Cancer Series no. 20).
- AIHW (Australian Institute of Health and Welfare), AACR (Australasian Association of Cancer Registries) & NHMRC National Breast Cancer Centre 1998. Breast cancer survival in Australian women 1982–1994. AIHW cat. no. CAN 4. Canberra: AIHW (Cancer Series no. 9).
- AIHW (Australian Institute of Health and Welfare) & DoHA (Department of Health and Ageing) (forthcoming). BreastScreen Australia Data Dictionary.
- BSANAC (BreastScreen Australia National Advisory Committee) & DHAC (Department of Health and Aged Care) 2000. BreastScreen Australia Evaluation Plan Phase II. Canberra: Commonwealth of Australia.
- BreastScreen Australia 1996. BreastScreen Australia statistical report 1996. Canberra: BreastScreen Australia.
- BreastScreen ACT 2000. BreastScreen ACT & SENSW annual statistical report 1998/1999. Canberra: BreastScreen ACT.
- BreastScreen Queensland 2000. Annual Statistical Report for 1997. Brisbane: BreastScreen Queensland.
- BreastScreen SA 1999. BreastScreen SA at 10 Years (incorporating the 1997 Statistical Report), Adelaide: BreastScreen South Australia.
- BreastScreen Victoria 2001. Annual statistical report, 1999. Carlton South: BreastScreen Victoria.
- BreastScreen WA 1999. BreastScreen WA statistical report 1996–1997. Perth: BreastScreen WA.
- Department of Health and Ageing (DoHA) (unpublished). Draft BreastScreen Australia Monitoring Plan.
- DHSH (Commonwealth Department of Human Services and Health) 1994. National Program for the Early Detection of Breast Cancer – minimum data set: for screening and assessment services. Canberra: Australia Government Publishing Service.

DPIE (Commonwealth Department of Primary Industries and Energy) & DSHS (Department of Human Services and Health) 1994. Rural, remote and metropolitan areas classification: 1991 Census edition. Canberra: Australian Government Publishing Service.

Day NE 1991. Screening for Breast Cancer. *British Medical Bulletin* 47:400-15.

Duffy SW, Tabar L, Fagerberg G, Gad A, Grontoft O, South MC & Day NE 1991. Breast Screening, prognostic facts and survival – results from the Swedish Two-Country Study. *British Journal of Cancer* 64:1133-1138.

Estoesta JV, Supramaniam R, Brassil AE & Taylor RJ 2000. BreastScreen New South Wales Ten Year Statistical Report: 1988-98. Sydney: BreastScreen NSW.

Feig SA. 1998. Decreased breast cancer mortality through mammographic screening: results in clinical trials. *Radiology* 167:659-665.

Fletcher SW, Black W, Harris R, Rimer V & Shapiro S 1993. Report of the International Workshop on Screening for Breast Cancer. *Journal of the National Cancer Institute* 85(20):1644-1656.

Kavanagh A, Amos AF & Marr GM 1999. The ascertainment and reporting of interval cancers within the BreastScreen Australia Program. Sydney: NHMRC National Breast Cancer Centre.

Kricker A & Jelfs P 1996. Breast cancer in Australian women 1921-1994. Canberra: AIHW (Cancer Series no. 7).

NBCC (National Breast Cancer Centre), AACR (Australasian Association of Cancer Registries), BSA (BreastScreen Australia), DHAC (Department of Health and Aged Care) & AIHW (Australian Institute of Health and Welfare) 2000. Ductal carcinoma in situ (DCIS). Canberra: AIHW (Cancer Monitoring Series no. 1).

NQMC (National Quality Management Committee of BreastScreen Australia) unpublished Draft National Accreditation Standards.