Appendix: Technical notes

Data sources (excluding the DCIS data set)

Australian Cancer Database (ACD), AIHW

The ACD contains unit record data, including persons' names, for every cancer diagnosed in Australia since 1 January 1982, excluding non-melanoma skin cancer. The state and territory cancer registries collect these data in their respective jurisdictions (and under their respective legislation) and supply the AIHW with an agreed subset of their data. This data set is standardised and deduplicated by the AIHW to become the ACD.

National Mortality Database (NMD), AIHW

The NMD contains de-identified unit record data for every death registered in Australia since 1 January 1968. The state and territory Registries of Births, Deaths and Marriages collect these data and supply them to the Australian Bureau of Statistics (ABS). The ABS uses the information on the death certificates to code the causes of death in ICD-10. By agreement with the registries the ABS then supplies these data, excluding persons' names, to the AIHW for compilation into the NMD.

National Death Index (NDI), AIHW

The NDI contains unit record data, including persons' names, for every death registered in Australia since 1 January 1980. The state and territory Registries of Births, Deaths and Marriages collect these data and supply the AIHW with an agreed data set, excluding cause of death data, which forms the NDI. The NDI is linked to the NMD, using the death certificate number as the linkage key, in order to transfer the causes of death from the NMD to the NDI.

The four-month rule

Before describing how the DCIS data set was constructed it is necessary to explain the 'four-month rule'. All of the Australian cancer registries use a business rule called the four-month rule, which in its simplest form states that if a woman is diagnosed with a DCIS and then, within four months, is diagnosed with an invasive breast cancer, the record of the DCIS shall be discarded. This rule is based on the consensus view that in such a situation the invasive breast cancer was almost certainly present at the time of the DCIS diagnosis but was not detected. The full version of the rule specifies how to deal with a DCIS in one breast and an invasive breast cancer in the other breast (and bilateral DCIS and/or bilateral invasive breast cancer), and how to deal with the various different histological types of invasive breast cancer that can arise. Unfortunately for this nationally-based project, different jurisdictions applied slightly different variations of this rule over the period of interest. Also, in some

circumstances in the Queensland cancer registry, the four-month time limit was extended to six months.

The DCIS data set

To be eligible for inclusion in the DCIS data set used in this project, a woman must not have had a diagnosis of invasive breast cancer prior to her diagnosis of DCIS.

The data set was constructed in the following manner.

1. Each of the state and territory cancer registries was asked to supply a data set of all its eligible DCIS cases (females only) from 1995 to 2005. South Australia and the Northern Territory were only able to supply data from 1997 and 1996 onwards, respectively. Each of the registries applied their own variant of the four-month rule to these data. For simplicity and national consistency, the AIHW applied an overriding four-month rule. The rule took no account of laterality of the DCIS and the invasive breast cancer and no account of the histological type of the invasive breast cancer—instead, the rule simply removed any DCIS record in which an invasive breast cancer was diagnosed in less than or equal to 121 days of a DCIS. This resulted in a data set of 13,792 records of DCIS.
2. The data set just constructed was deduplicated (that is, linked with itself) and 21 duplicate pairs (that is, women with DCIS records in two jurisdictions) were detected. Within each pair, the record with the earliest diagnosis date was retained and the other was removed, leaving 13,771 records. Of these, there were 689 known cases of invasive breast cancer and 557 known deaths ('known' meaning that the information was recorded on the original data sets supplied by the cancer registries).
3. The data set was then linked to the NDI. This resulted in the identification of 590 women in the DCIS data set who were dead but had not been recorded as dead in the original data sets from the registries. The DCIS data set was updated with the fact and date of death for each of these women, bringing the total number of deaths in the DCIS data set to 1,147.
4. In order to detect instances of a woman being diagnosed with DCIS in one jurisdiction and invasive breast cancer in a different jurisdiction, the DCIS data set was linked to the ACD. This resulted in a further 39 cases of invasive breast cancer being discovered. However, for 22 of these women, the invasive breast cancer had been diagnosed either before or within four months of the DCIS diagnosis. Consequently, these 22 women were removed from the data set and the remaining 17 cases of invasive breast cancer were recorded. Of the 22 women deleted, three were dead. Thus, the resulting data set consisted of 13,749 records that included 706 cases of invasive breast cancer and 1,144 deaths.
5. Finally, the relevant cancer registries were asked to supply the tumour size and nodal status for the 17 additional cases of invasive breast cancer that were uncovered in the preceding step.

Caveats

Consider a woman whose DCIS diagnosis occurred between September and December of 2005 who was diagnosed with invasive breast cancer in 2006 within four months of her DCIS diagnosis. Application of the four-month rule would result in her being removed from the

DCIS data set. Since the follow-up period for this study did not extend into 2006 there was no way of identifying such women and it was decided to keep all of them in the DCIS data set. The small number that would have been removed if 2006 data were available would not have appreciably altered the study's findings.

As already mentioned, in some circumstances the Queensland cancer registry applied a 'six-month rule' instead of the four-month rule. Therefore, there were some women who ideally should have been in the DCIS data set used for this study but were not. Based on the project team's analysis of the national data set and discussion with the Queensland cancer registry, it is felt that there would have been very few cases excluded in this way and that their inclusion would not have altered the study's findings.

Some extra information regarding the Kaplan-Meier estimates of probability of invasive breast cancer

The Kaplan-Meier product limit technique was used to calculate the probabilities of invasive breast cancer (Chapter 3). Follow-up commenced at DCIS diagnosis. The event of interest was diagnosis of invasive breast cancer. Censoring occurred at death or 31 December 2005 (i.e. the end of the study period), whichever occurred first. The shortest period elapsing to diagnosis of invasive breast cancer was 123 days. (Given the four-month rule, the shortest possible such period was 122 days.) By comparison, the longest elapsed time was 3,806 days (over 10 years). Of the 13,749 women in the DCIS cohort, 12,038 (87.6%) were censored at the end of the study period and 1,005 (7.3%) at death (Table A.1).

Table A.1: Number of women in various categories in the DCIS cohort

	Invasive status		
Death status	Diagnosed with invasive breast cancer during study period	Not diagnosed with invasive breast cancer during study period	Total
Died during study period	139	1,005	1,144
Did not die during study period	567	12,038	12,605
Total	**706**	**13,043**	**13,749**

The type of censoring varied with age. For women aged less than 40 years, 40–49 years, 50–59 years, 60–69 years, 70–79 years and 80 years or over, end-of-study censoring rates were 88.5%, 92.8%, 93.2%, 88.0%, 75.9% and 49.3% respectively, whereas death censoring rates were 2.5%, 1.4%, 2.8%, 7.0% 18.3% and 46.1% respectively.

Details of the test used to detect a difference between observed and expected tumour size and nodal status (Chapter 5)

To keep matters concrete, the following discussion will focus on the number of tumours of size 0–10 millimetres in women aged 50 to 59 years (see Tables 5.1–5.4). The number, X, of tumours of this size in this age group was treated as a binomial random variable. The

'number of trials' is n = 125, which was derived from the fact that there were a total of 125 cases of invasive breast cancer with known tumour size in this age group. The 'probability of success' is unknown but taken to be p = 587/2,302 = 0.255 from the observed Australia-wide distribution in 1997 as shown in Table 5.1. Therefore the expected value of X is $\mu = np = 125 \times 0.255 = 31.9$, as shown in Table 5.3, and the standard deviation of X is $\sigma = \sqrt{np(1-p)} = \sqrt{(31.9 \times 0.745)} = \sqrt{23.7} = 4.87$ (not shown in the tables).

Using conventional practice it was assumed that X is approximately normally distributed because $n \geq 30$ and p is not extremely close to either 0 or 1. Therefore the statistic $Z = (X-\mu)/\sigma$ is approximately a standard normal random variable.

The tests employed were one-sided tests because it was conjectured beforehand that there would be more smaller tumours than expected and fewer large ones. This was quantified by conjecturing more smaller tumours for the two smallest size categories (0–10 mm and 11–15 mm) and fewer large ones for the other three size categories. A 5% level of significance was used. Therefore the alternative hypothesis was accepted for the two smallest size categories when $z \geq 1.645$ and for the three largest size categories when $z \leq -1.645$.

Continuing with the example, the observed value of X was x = 52 (Table 5.3). Therefore the observed value of Z was $z = (52-31.9)/4.87 = 4.13$, which is statistically significantly higher than would be expected.

The same method was used to detect a difference in the number of tumours of a given nodal status in a given age group. The alternative hypothesis was that there would be more node-negative cancers (equivalently, fewer node-positive ones).

Glossary

Age-standardised rate: a method of removing the influence of age when comparing populations with different age structures. This is usually necessary because the rates of many diseases vary strongly (usually increasing) with age. The age structures of the different populations are converted to the same 'standard' structure, which allows comparison of disease rates.

Cancer (malignant neoplasm): a large range of diseases in which some of the body's cells become defective, and begin to multiply out of control. These cells can invade and damage the area around them, and can also spread to other parts of the body to cause further damage.

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 lies.

Ductal carcinoma in situ (DCIS): a lesion in the breast in which cancerous cells have not infiltrated beyond the duct walls to surrounding tissue.

Malignant: abnormal changes consistent with cancer.

References

Acronyms used in references

AIHW	Australian Institute of Health and Welfare
BSA	BreastScreen Australia
IARC	International Agency for Research on Cancer
NBCC	National Breast Cancer Centre
WHO	World Health Organization

References

AIHW 2008. BreastScreen Australia monitoring report 2004–2005. Cancer series no. 42. Cat. no. CAN 37. Canberra: AIHW.

AIHW & NBCC 2001. Breast cancer size and nodal status. Cancer monitoring series no. 2. Canberra: AIHW.

AIHW & NBCC 2007. Breast cancer survival by size and nodal status in Australia. Cancer series no. 39. Cat. no. CAN 34. Canberra: AIHW.

Burstein HJ, Polyak K, Wong JS, Lester SC & Kaelin CM 2004. Ductal carcinoma in situ of the breast. New England Journal of Medicine 350:1430–41.

Claus EB, Stowe M, Carter D & Holford T 2003. The risk of contralateral breast cancer among women diagnosed with ductal and lobular breast carcinoma in situ: data from the Connecticut Tumor Registry. Breast 12:451–56.

Habel LA, Moe RE, Daling JR, Holte S, Rossing MA & Weiss NS 1997. Risk of contralateral breast cancer among women with carcinoma in situ of the breast. Annals of Surgery 225:69–75.

Innos K & Horn-Rors PL 2008. Risk of second primary breast cancers among women with ductal carcinoma in situ of the breast. Breast Cancer Research and Treatment 111:531–40.

Kricker A, Goumas C & Armstrong B 2004. Ductal carcinoma in situ of the breast, a population-based study of epidemiology and pathology. British Journal of Cancer 90:1382–85.

Li CI, Malone KE, Saltzman BS & Daling JR 2006. Risk of invasive breast carcinoma among women diagnosed with ductal carcinoma in situ and lobular carcinoma in situ, 1988–2001. Cancer 106:2104–12.

Luke C, Priest K & Roder D 2006. Changes in incidence of in situ and invasive breast cancer by histology type following mammography screening. Asian Pacific Journal of Cancer Prevention 7:69–74.

Patani N, Cutuli B & Mokbel K 2008. Current management of DCIS: a review. Breast Cancer Research and Treatment 111:1–10.

Wärnberg F, Yuen J & Holmberg L 2000. Risk of subsequent invasive breast cancer after breast carcinoma in situ. The Lancet 355:724–25.

WHO & IARC 2002. Breast cancer screening. IARC handbooks of cancer prevention vol. 7. Lyon: IARC Press.

List of tables

List of figures