Effect of improved data collection on breast cancer incidence and survival: reconciliation of a registry with a clinical database

Anne Stotter, Nicola Bright, Paul B Silcocks, Johannes L Botha

Glenfield Hospital
NHS Trust,
Leicester LE3 9QP
Anne Stotter
consultant surgeon

Trent Cancer
Registry, Weston
Park Hospital NHS
Trust, Sheffield
S10 2SJ
Nicola Bright
medical statistician
Paul B Silcocks
*assistant director
(research and
intelligence)*
Johannes L Botha
director

Correspondence to:
A Stotter
anne.stotter@
glenfield-tr.trent.
nhs.uk

In 1998 a project was undertaken to improve cancer registration in the Trent region by establishing a direct link between Trent Cancer Registry and a breast cancer clinical database for Leicestershire patients at Glenfield Hospital. This provided an opportunity to study registrations for 1997 to determine how many and who were missed. We then estimated the effect of improved ascertainment on incidence and survival.

Methods and results

Throughout 1998 Trent Cancer Registry sent registrations for 1997 to Glenfield Hospital for comparison. We investigated three main issues.

● Differences in recorded date of diagnosis, including identification of an earlier date of diagnosis for patients registered as diagnosed at the time of death (that is, where diagnosis is based solely on death certificate information)

● Characteristics of those who were missed

● Effects of these factors on the apparent incidence and survival of breast cancer patients.

Using Leicestershire breast cancer registrations for 1993 on the registry database, we fitted a Cox model[1] to the age specific survival and used it to predict survival for the 1997 cohort.

At the cancer registry 535 breast cancers were registered for 1997. On the Glenfield database the date of diagnosis was a median of 26 days earlier than the date of registration; 70 registrations were assigned to a previous year. By the end of 1998, 134 patients listed on the Glenfield database were still not registered. Thus, 599 diagnoses were finally identified for 1997, a 12% increase.

The median age of those registered was 58 years, compared with 74 for those missed (the median age of those who underwent surgery was 56 compared with 77 for those who did not). All 12 patients receiving private care were missed by the registry. Of 62 registrations based solely on death certificate information, 25 had an earlier date of diagnosis on the Glenfield database; they had a median survival of 11 months.

Using the model, we predicted five year survival to be 62% in the 535 registered cases and 59% in all 599 cases (figure).

Comment

Improved ascertainment of breast cancer registrations apparently increased the incidence of cancer and, counter intuitively, seemed to reduce survival. This was because the missed cases were not typical of the cohort as a whole but comprised a subset with a lower life expectancy.

The main source of notifications for the registry was through the hospital patient administration system—that is, inpatients and day cases only. Patients managed as outpatients (such as those treated with tamoxifen
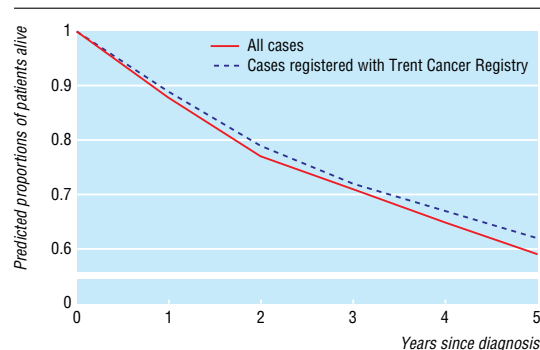
Predicted survival of Leicestershire patients with breast cancer diagnosed in 1997

only) were likely to be missed. The addition of the missing cases reduced the overall survival because they were predominantly older women, with a shorter survival.

Additionally, the registry is notified of the death of all registered patients and of those for whom cancer is mentioned as a cause of death. Registrations based solely on death certificate information are conventionally excluded from survival statistics, and the Glenfield database now allows a search for an earlier diagnosis date. The short median survival of 11 months for these patients contributes to a reduction in overall survival.

Standardisation of the definition of the diagnosis date in the registry to that used in the Glenfield database will result in nearly one month's improvement in the apparent survival of patients with breast cancer. This small increase will not, however, outweigh the other factors.

National cancer statistics are used as evidence of "black spots" in cancer incidence and survival.[2][3] It is crucial that those who use them recognise that statistics are influenced by many factors besides cancer epidemiology and treatment, in particular the details of how the data are collected. This becomes increasingly important when attempts are made to compare different centres and even countries.[4]

1 Cox DR, Regression models and life-tables (with discussion). *J R Stat Soc B* 1972;34:187-220.
2 *Cancer statistics—registrations of cancer diagnosed in 1992, England and Wales.* London: Office for National Statistics, 1998.
3 Coleman MP, Babb P, Mayer D, Quinn MJ, Sloggett A. *Cancer survival trends in England and Wales, 1971-1995: deprivation and NHS region.* London: Office for National Statistics, 1999.
4 Quinn MJ, Martinez-Garcia C, Berrino F, EUROCARE Working Group. Variations in survival from breast cancer in Europe by age and country, 1978-1989. *Eur J Cancer* 1998;34:2204-11.

(Accepted 25 April 2000)