ORIGINAL ARTICLE

Taylor&Francis healthsciences

Validity and Representativity in the Danish Breast Cancer Cooperative Group

A Study on Protocol Allocation and Data Validity from One County to a Multi-centre Database

Anni R. Jensen, Hans H. Storm, Susanne Møller and Jens Overgaard

From the Cancer Documentation, Department of Experimental Clinical Oncology, (A.R. Jensen, J. Overgaard), The Danish Cancer Society Aarhus University Hospital, Denmark, the Centre for Cancer Documentation, Department of Prevention and Documentation (H.H. Storm), Danish Cancer Society, Copenhagen, Denmark and the Danish Breast Cancer Cooperative Group (S. Møller), National University Hospital (Rigshospitalet), Copenhagen, Denmark

Correspondence to: Anni R. Jensen, Department of Experimental Clinical Oncology, Aarhus University Hospital, Norrebrogade 44, Building 5, DK-8000 Aarhus C, Denmark. Tel: +45 89 492 620. Fax: +45 86 197 109. E-mail: anni@oncology.dk

Acta Oncologica Vol. 42, No. 3, pp. 179–185, 2003

Population-based cancer registries collect basic information on incident cases, whereas clinical databases hold a broad range of variables that make them attractive sources in epidemiological research. Quality, completeness and representativeness of the clinical database in the Danish Breast Cancer Cooperative group (DBCG) were compared with those in a complete population-based database holding detailed information on patients with breast cancer. The study included 1 765 patients diagnosed in Aarhus County between 1983 and 1989. Data on tumour characteristics were of high quality. Fifty percent of patients were included in DBCG trials and a total of 83% were notified in the DBCG register, the rest having significantly poorer stage distribution and survival probabilities. However, for patients younger than 70 years of age, the difference in overall survival between all patients and those notified in the DBCG register was only 2%. Follow-up was incomplete in patients treated outside the programme, and this had a major impact on recurrence-free survival. The DBCG database holds high-quality data, but as a population-based database, it is neither comprehensive nor representative.

Received 13 September 2002
Accepted 6 February 2003

The Danish Breast Cancer Cooperative Group (DBCG) was established in 1977 as initiator and organizer of nation-wide clinical guidelines and prospective adjuvant trials on patients with primary breast cancer (1). In addition, the group evaluates breast cancer treatment programmes based on follow-up for 10 years and also seeks to collect clinical data on patients not included in the programme. The database holds information on primary surgical procedures, histopathology, adjuvant therapy and clinical follow-up. There have been numerous publications based on these trials, and, furthermore, data from randomized trials have been analysed in a variety of other contexts (2). However, high-quality data are essential and it is important to address the quality of data, the representativeness and the potential risk and magnitude of selection bias. These issues are especially important when a register that is regarded as a population-based is used for studies. The aim of the present study was to validate the data quality and the completeness and generalizability of the cases notified in the DBCG register.

MATERIAL AND METHODS

Study population

The study included females with breast cancer living in the county of Aarhus and diagnosed from 1 January 1983 to 31 December1989, a time period within which the DBCG-82 trial was held. Incident cases were identified by linkage between the Danish Cancer Registry, the DBCG register and the register at the local Oncology Department. The Danish Cancer Registry is well known for its comprehensive and high-validity data and is fully described elsewhere (3). All linkages were done on the basis of the unique central population registration number, assigned to all Danes in and after 1968. A total of 2 000 patients with invasive breast cancer were identified from the combined registers. The

 ISSN 0284-186X
DOI: 10.1080/02841860310000737

clinical records of these patients were sought from hospitals and general practitioners, and the established clinical database holding data from medical records was defined as the gold standard (4). In the present study we excluded patients with a previous history of invasive cancer, patients known from death certificates only and patients without a histopathological diagnosis. A total of 1 765 patients with primary invasive breast carcinoma were included in the study population. Clinical records were found for all patients, with the exception of 16 ($< 1\%$).

The DBCG

The DBCG register holds detailed descriptions of tumour localization, size, histopathological characteristics, nodal involvement and surgical procedures. The observational programme for patients included in the DBCG trials includes information on adjuvant treatment and 10-years' follow-up, first recurrence, development of a new primary cancer or death, whichever comes first. Date and reasons for off-study are reported on a specific off-study record. If recurrence is the reason, the site of the recurrence is indicated.

The DBCG-82 programme consisted of four different protocols to which patients were allocated according to risk group, menopausal status and whether they were technically fit for breast-conserving surgery (Fig. 1). The low-risk group included patients with tumours ≤ 50 mm, no invasion of deep resection line/skin and no axillary metastasis. The DBCG trial for the low-risk group was designed as a prospective follow-up study without randomization. Patients with tumours greater than 50 mm, microscopic tumour invasion of either the skin or deep resection line or with axillary lymph node metastasis were allocated to the high-risk group. Patients in the high-risk groups and patients eligible for lumpectomy were randomized to different treatment modalities within each group. The protocols remained almost unchanged during the study period. Patients were excluded from the trials if they were older than 69 years, had bilateral tumours or distant metastases, if tumours were inflammatory or fixed to the skin or deep muscle, if surgery was not performed in accordance with the guidelines, if they had a history of previous invasive cancer or if there were any concomitant medical diseases.

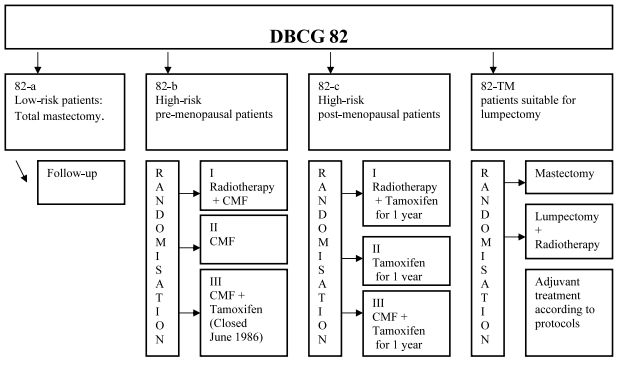

Fig. 1. Design of the DBCG-82 trials CMF regime: cyclophosphamide, methrotrexat, fluorouracil. Tamoxifen: 30 mg daily. CMF = cyclophosphamide, methrotrexat, fluorouracil.

Definitions

The terminology 'notified in DBCG' is used for patients notified in the DBCG register and the terminology 'unknown' is used for the remainder of the patients. For the group of patients notified in the DBCG we made a further distinction between patients included in the DBCG trial 'in DBCG trial' and those 'notified-not in trial'.

Data analysis and follow-up

We compared the DBCG register with the clinical database. Records were linked by means of the unique personal identification number allocated to each citizen in Denmark, yielding one record per person. If there were discrepancies of more than half a year in date of diagnosis, the record from the DBCG was excluded from further validation.

Registries from the national health authorities facilitated the following-up of patients. Patients were followed from the date of surgery for primary breast cancer to 31 December 2000. The median potential follow-up time (time from entry until date of assessment) was 14.4 years (range 11.0–18.0 years). Clinical status, localization and date of first recurrence were traced from the medical records. A locoregional recurrence was defined as a recurrence located to the skin of the ipsilateral chest wall, infra/supraclavicular, axillary or parasternal lymph nodes. All other locations were considered distant. All sites noted within 30 days of the first recurrence were considered part of the first recurrence (5).

Overall survival was calculated as the interval from date of surgery/biopsy to death or end of follow-up. Recurrence-free interval was defined as time without recurrence or cancer in the opposite breast. This implies that patients are censored at death and all patients with distant metastases at diagnosis have an event at the beginning of the Kaplan–Meier curve. Disease-free survival is the length of time without recurrence, cancer in the opposite breast or death. Actuarial rates were estimated using the Kaplan–Meier method. The 95% confidence intervals (CI) were calculated as asymmetric intervals. Comparisons between groups were made using the Mann–Whitney and the Kruskal–Wallis tests and Spearman's test was used for assessing trend. Level of significance was 5% and all p-values were two-tailed.

RESULTS

Fifty percent (885/1 765) patients were included in DBCG trials and an additional 33% (578/1 765) were to some extent notified in the DBCG register (Table 1). Eleven percent

Table 1

Notification in the DBCG register and protocol allocation in DBCG trials according to year of surgery/biopsy

	In DBCG trial*	Notified in DBCG not in trial	Subtotal in DBCG register	Unknown in DBCG#	No exclusion criteria not in trial*	Total
1983	136 (61)	65 (29)	201 (90)	22 (10)	9 (4)	223 (100)
1984	131 (54)	74 (31)	205 (85)	37 (15)	21 (9)	242 (100)
1985	121 (50)	88 (36)	209 (86)	34 (14)	23 (10)	243 (100)
1986	137 (53)	80 (31)	217 (83)	43 (17)	28 (11)	260 (100)
1987	114 (45)	81 (32)	195 (76)	61 (24)	41 (16)	256 (100)
1988	140 (51)	76 (28)	216 (79)	58 (21)	32 (12)	274 (100)
1989	106 (40)	115 (43)	221 (83)	46 (17)	57 (21)	267 (100)
	885 (50)	579 (33)	1464 (83)	301 (17)	211 (12)	1765 (100)

Percentages of study cohort in parentheses.
* Time trend $p < 0.01$.
# Time trend not significant.

(137/1 232) under the age of 70 were unknown to the DBCG and for all ages 17% (301/1 765) were unknown (Table 2). A comparison was made between patients 'notified-not' in trials and those unknown to the DBCG (Table 2). The extent of axillary surgery in terms of lymph nodes removed was significantly higher in those 'notified-not' in trials. For patients below 70 years, the difference in stage distribution was of borderline significance, whereas for the older age group stage distribution was significantly less favourable for patients unknown to the DBCG. There were significant differences in overall survival between patients notified and those unknown to the DBCG, in 10 years 51% (48–53) vs. 23% (19–28), whereas the whole patient population had a 10-year survival rate of 46% (44–48) (Fig. 2). However, the difference between the whole cohort and patients notified in the DBCG was small for patients younger than 70 years where the 10-year survival was 56% (53–59) vs. 58% (55–61).

The main reason for not entering the DBCG trials was age above 70 year. Among patients younger than 70 not fulfilling any exclusion criteria, 17% (211/1 233) were not included in the DBCG trials. We could not find data from the medical records on why these patients were not entered in a trial. We found a significant trend towards decrease in time for patients entering trials, and a corresponding increase in eligible patients not allocated to the protocol. An analysis of survival did not show any difference in survival between patients allocated to a protocol and those eligible but not included, either in the low- or high-risk group. The regimens on adjuvant treatment in the randomized trials were highly concordant with the actual treatment given, as we found protocol violations in only 1.5% of all high-risk patients (5/346) included in the DBCG trials.

Data on tumour characteristics in the DBCG register were of high quality, being more than 99% accurate for all

Table 2

Patient characteristics according to whether they were included in the DBCG-82 trial, notified in the DBCG register or unknown to the DBCG

	In DBCG trial	< 70 years, not in trial			> = 70 years, not in trial		
		Notified DBCG	Unknown DBCG	p-value	Notified DBCG	Unknown DBCG	p-value
Number of patients	885 (100)	211[a] (100)	137 (100)		368 (100)	164 (100)	
Age							
Median (range)		52.2 (28–70)	55.3 (29–70)	NS	76.4 (70–93)	80.6 (70–95)	< 0.001
Tumour size							
Median (range)		20 (5–150)	20 (5–100)	NS	20 (2–100)	22 (5–120)	NS
Lymph nodes removed							
Median (range)		5 (0–23)	4 (0–16)	< 0.001	4 (0–19)	0 (0–19)	< 0.001
Stage (UICC)				< 0.05			< 0.001
I	392 (44)	46 (22)	27 (20)		97 (26)	7 (4)	
II	465 (53)	91 (43)	39 (28)		167 (45)	49 (30)	
III	27 (3)	15 (7)	24 (18)		54 (15)	39 (24)	
IV	1 (0)	49 (23)	39 (28)		11 (3)	28 (17)	
Unclassified	0	10 (5)	8 (6)		39 (11)	41 (25)	

[a] Includes a group of patients not included in a randomized trial but treated in accordance with standard trial arms, and a group with advanced disease not suitable for trial. Percentages in parentheses.

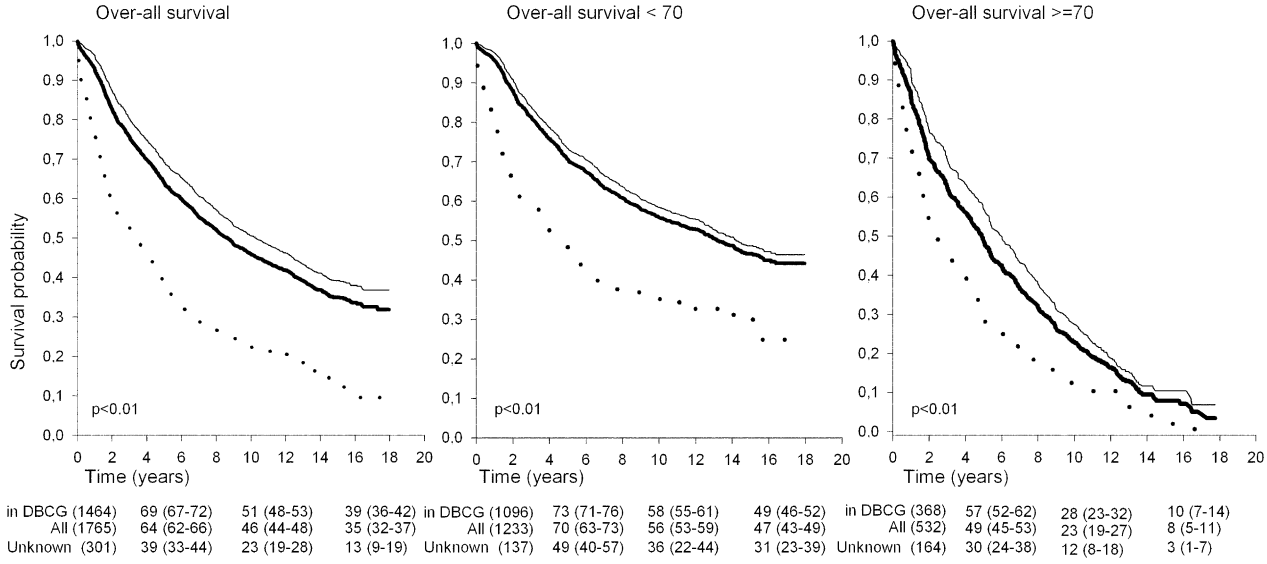

Fig. 2. Kaplan–Meier estimates of overall survival in the total cohort and among women notified in the DBCG register and those unknown to the DBCG. Notified in DBCG register: — All patients. · · · Unknown in the DBCG. Survival rate and 95% confidence intervals at 5, 10 and 15 years; p-value between patients notified and all patients.

variables (Table 3). However, for several variables there were up to 10% of missing values in comparison to what was obtainable from the clinical records. Concerning dates, both the date of biopsy and date of surgery were correctly notified in more than 90% of cases but the date of biopsy was missing in 17% of cases.

The DBCG register shows high concordance in tumour laterality, being correctly notified in more than 95% of

Table 3

Completeness and correctness of data on tumour characteristics in the DBCG register for 1 464 patients notified in the register. Percentages in parentheses

	Clinical data accessible	Completeness in DBCG	Correctness of notification in DBCG
Menopausal status	1 464	1 455 (99)	1 454 (100)
Date of biopsy	1 120	934 (83)	852 (91)*
Date of final surgery	1 464	1 457 (99)	1 344 (92)**
Number of lymph nodes removed	1 456	1 373 (94)	1 366 (99)
Number of positive lymph nodes	1 364	1 313 (92)	1 310 (100)
Macroscopic multicentric tumour	1 344	1 202 (89)	1 202 (100)
Residual tumour in cavity	1 293	1 164 (90)	1 158 (99)
Tumour size (largest diameter)	1 423	1 344 (94)	1 344 (100)
Invasion of skin	1 424	1 285 (90)	1 284 (100)
Invasion in profound resection line	1 396	1 274 (91)	1 273 (100)
Invasion in side resection line	1 359	1 228 (90)	1 227 (100)
Histopathological type	1 464	1 317 (90)	1 317 (100)
Tubulus formation	1 221	1 125 (92)	1 125 (100)
Mitosis	1 222	1 119 (92)	1 117 (100)
Nucleus pleomorphism	1 244	1 122 (90)	1 122 (100)
Mb. Paget of the nipple	1 345	1 227 (91)	1 227 (100)
+Carcinoma in situ	1 402	1 272 (91)	1 271 (100)
Microscopic multicentric tumour	1 389	1 260 (91)	1 258 (100)
Lymph nodes; perinodal invasion	1 388	1 228 (88)	1 228 (100)
Invasion in nerve	1 258	1 176 (93)	1 176 (100)
Vascular invasion	1 255	1 228 (98)	1 228 (100)
Lymph vessel invasion	1 239	1 151 (93)	1 151 (100)
Side localisation (right/left)	1 464	1 360 (93)	1 301 (96)
Tumour location within the breast	1 464	1 256 (86)	987 (73)

* Date of first biopsy earlier in 76 and later in 6 cases.
** Date of final surgery was earlier than notified in 85 patients.

patients, but the results are less precise for tumour location within the breast.

For patients included in the DBCG trials, completeness of reporting recurrences was 93% (383/411), whereas for all patients notified in the DBCG the reporting rate was 64% (227/633) (Table 4). Concerning the probability of recurrence, there were major differences between estimates based on the DBCG register and the data set from the medical records, irrespective of age group (Fig. 3). The observed selection in the reporting of recurrences had an impact on the estimate of disease-free survival for patients older than 69 years but not for the younger group (Fig. 5). Time from recurrence to death also differed according to whether recurrence was reported or not. Median survival time from recurrence was 1.9 (1.7–2.2) years for all patients with a recurrence, 2.5 (2.2–2.7) years for patients with first recurrence reported to the DBCG and 1.4 (0.9-1.8) years for patients where the first recurrence was not reported.

Site of first recurrence in terms of locoregional or distant metastases were correct in 95% (365/383) for patients allocated to the DBCG trials as well as for all patients notified to the DBCG register (386/406).

DISCUSSION

The DBCG register has in previous studies been considered a population-based register with a completeness rate of 90–95% of all incident cases in patients younger than age 70 years (6, 7). A recent study linking the Danish National Cancer Registry and the DBCG register concluded that the DBCG can not replace the National Cancer Registry as a reliable source for incident breast cancer cases, as 18% were missing in the DBCG register (8). This result is in agreement with our study where 17% of cases were found unknown to the DBCG register. The DBCG is a nationwide register but notification is not mandatory and there are no routines to ensure notification either for inhospital patients or for patients treated outside the hospital. However, with the intention of the DBCG to include all cases treated in hospitals, it may be regarded as a population-based register with incomplete coverage (9). Analysing the DBCG register as a population-based register introduces selection bias, as the group of patients with the worst survival probability is unknown. However, for patients below the age of 70, the DBCG register will only have a minor effect on outcome measurements in terms of overall survival.

Patients are withdrawn from follow-up if they develop a new primary cancer. This perhaps explains the considerable underreporting of recurrences but also the small impact this has on disease-free survival. In the DBCG studies, disease free-survival is defined as survival without recurrence, cancer in the opposite breast or other malignant disease (10). This will decrease the differences between our estimate and the DBCG estimate, and disease-free survival thereby seems to be a safe outcome measurement for patients younger than 70 years of age (Fig. 3). Data on first recurrence were missing for 34% (227/633) of patients notified (Table 4), which was not surprising, as the DBCG is intended actively to collect clinical follow-up on patients included in DBCG trials only and for these patients 93% (383/411) were reported. The observed difference in median survival time after first recurrence, according to whether recurrences were reported to the DBCG or not, could have been introduced by fewer hospitalizations in oncology wards for patients whose recurrences were not reported

Table 4

Verification of registration of first recurrence in the DBCG register vs. clinical records

Recurrence in DBCG	Recurrence in medical records			
	In the first 10 years			In follow-up time
	+Recurrence	−Recurrence	Total	+Recurrence
In DBCG trial				
+Recurrence	383 (93)	6* (1)	389 (44)	386 (87)
−Recurrence	28 (7)	468 (99)	496 (56)	60 (23)
Total	411 (100)	474 (100)	885 (100)	446 (100)
Notified in DBCG				
+Recurrence	406 (64)	6* (1)	412 (28)	409 (60)
−Recurrence	227 (36)	825 (99)	1 052 (72)	268 (40)
Total	633 (100)	831 (100)	1 464 (100)	677 (100)
Total cohort				
+Recurrence	406 (54)	6* (1)	412 (23)	409 (51)
−Recurrence	347 (46)	1 006 (99)	1 353 (77)	396 (49)
Total	753 (100)	1 012 (100)	1 765 (100)	805 (100)

Recurrences reported to the DBCG register for patients included in the DBCG-82 trials, patients notified in the DBCG register and the whole cohort vs. information from the medical records.
A 10-year follow-up time is intended in the DBCG programme. Potential follow-up time was 14.4 years.

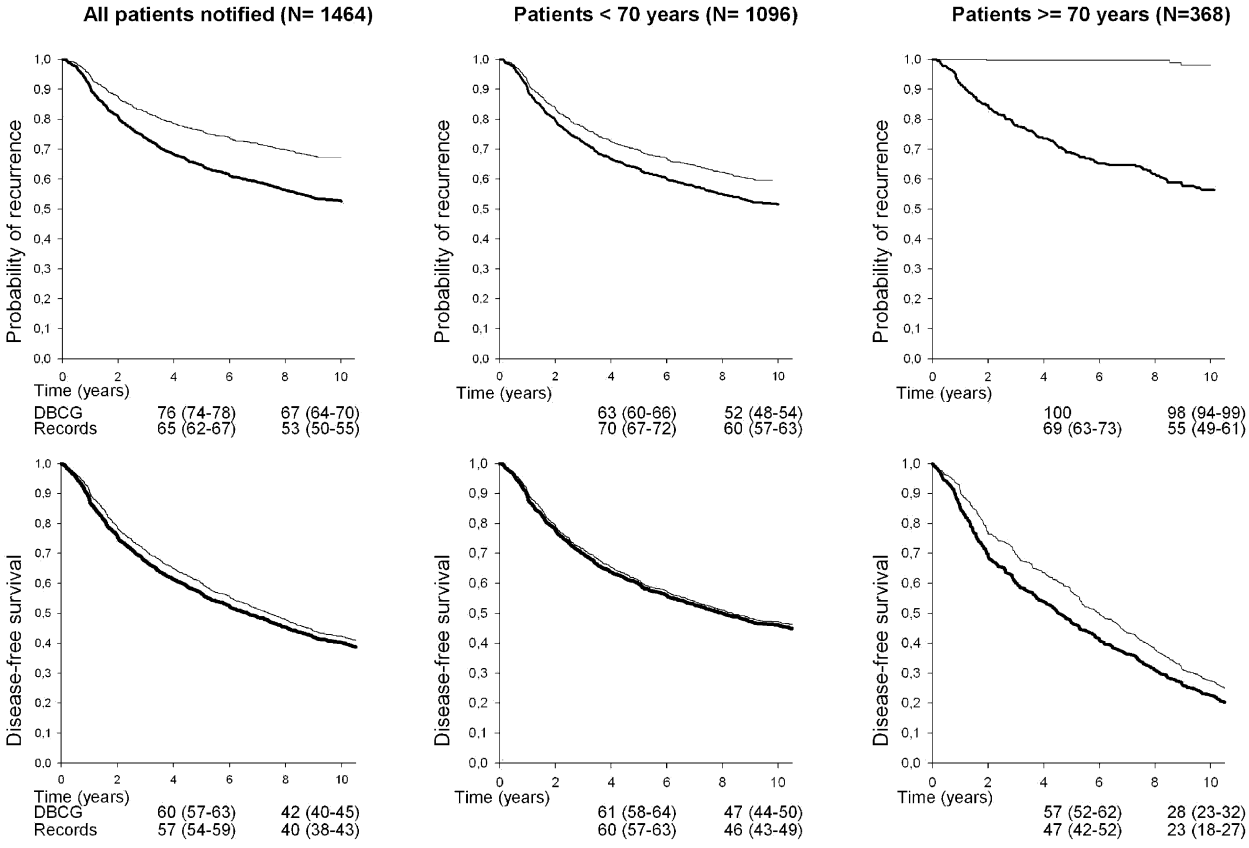

Fig. 3. Kaplan–Meier estimates of probability of recurrence and disease-free survival for 1 464 patients notified in the DBCG register according to whether information on recurrence is from the DBCG register or from the medical records. — Data from the medical records. — Data from the DBCG register. Rate and 95% confidence intervals at 5 and 10 years.

owing to old age, concomitant disease or more severe dissemination of metastases.

For the 885 patients who were allocated to the DBCG trials, information on first recurrence was missing for 28 patients only. This corresponds to a former quality control of recurrences in the DBCG 82b and 82c trials where follow-up data of high validity were found (11).

The proportion of patients included in the DBCG trials gradually decreased over time. We found a corresponding increase in the fraction of patients 'notified-not' in the trial. This indicates that physicians being less aware of the DBCG did not cause the decrease in inclusion to the trial. The decrease could be due to a demand for newer treatment modalities than those offered in a 7-year-old protocol, but analysis of treatment given to patients outside the protocols does not disclose a tendency towards more or other types of treatment. The trend also could be a result of changes in ethical rules on informed consent in Denmark during the study period, with an increasing number of patients showing unwillingness to undergo randomization (12).

The quality of the data for basic characteristics such as tumour laterality, tumour size, histopathology and number of lymph nodes examined was high as well as being reliable as a source of secondary data. The interpretation of information on tumour location within the breast was more complicated. Whenever a tumour was located in the borderline between two areas, it was assigned to one of the two adjacent areas by randomization according to date of birth (equal/unequal). The prognostic significance of tumour location within the breast has previously been analysed in epidemiological studies using the DBCG register but this variable is not precise and can only be used with caution (13). However, the misclassification introduced is non-differential and will tend to diminish risk estimates related to tumour localization (medial/lateral).

We defined the date of biopsy as being the date of first biopsy and the date of surgery as the date of final surgery. In cases with discrepancies, the time difference between dates at operation were small and of little or no importance when assessing survival from date of surgery. The interpretation of small time periods such as delay from biopsy to surgery require precise dates and are therefore dubious if based on data from the DBCG register.

CONCLUSION

We found a completeness rate of 83% in the DBCG register but a lack of representativeness, as those unknown to the registry comprised a subset with a lower life expectancy. However, for patients younger than 70 years, there were no significant differences in overall survival between the population-based cohort and the patients notified in the DBCG register. For patients notified in the DBCG register, follow-up reports on recurrences were only valid for those entered in clinical trials and recurrence free-survival was not a valid endpoint. This did not influence disease-free survival, which seems a valid endpoint for patients beyond 70 years of age.

In general, the DBCG register holds high-quality data concerning tumour characteristics and can be regarded as a valid source of secondary data for other studies. When analysing a population-based register, this must be done with caution because of the risk of selection bias. This study deals with patients from one county in one time period and a continuous quality control is recommended.

ACKNOWLEDGEMENTS

The study was supported by grants from the Danish Cancer Society, the Danish Medical Research Council and Aarhus University.

REFERENCES

1. Andersen KW, Mouridsen HT. Danish Breast Cancer Cooperative Group (DBCG). A description of the register of the nation-wide programme for primary breast cancer. Acta Oncol 1988; 27: 627–47.
2. DBCG. Copenhagen: DBCG Danish Breast Cancer Cooperative Group 1977–1997; 1998. p. 5–91.
3. Storm HH, Michelsen EV, Clemmensen IH, Pihl J. The Danish Cancer Registry—history, content, quality and use. Dan Med Bull 1997; 44: 535–9.
4. Jensen AR, Overgaard J, Storm HH. Validity of breast cancer in the Danish Cancer Registry. A study based on clinical records from one county in Denmark. Eur J Cancer Prev 2002; 11: 359–64.
5. Kamby C. The pattern of metastases in human breast cancer: methodological aspects and influence of prognostic factors. Cancer Treat Rev 1990; 17: 37–61.
6. Nielsen JC, Andersen E, Zedeler K. Breast cancer in the county of Viborg during 1983–1989. A quality control study. Ugeskr Laeger 1994; 156: 2737–41.
7. Grabau DA, Jensen MB, Blichert-Toft M, et al. The importance of surgery and accurate axillary staging for survival in breast cancer. Eur J Surg Oncol 1998; 24: 499–507.
8. Rostgaard K, Holst H, Mouridsen HT, Lynge E. Do clinical databases render population-based cancer registers obsolete? The example of breast cancer in Denmark. Cancer Causes Control 2000; 11: 669–74.
9. Olsen J, Basso O, Sorensen HT. What is a population-based registry? Scand J Public Health 1999; 27: 78.
10. Overgaard M, Hansen PS, Overgaard J, et al. Postoperative radiotherapy in high-risk premenopausal women with breast cancer who receive adjuvant chemotherapy. Danish Breast Cancer Cooperative Group 82b Trial. N Engl J Med 1997; 337: 949–55.
11. Hansen PS, Andersen E, Andersen KW, Mouridsen HT. Quality control of end results in a Danish adjuvant breast cancer multi-center study. Acta Oncol 1997; 36: 711–4.
12. Blichert-Toft M, Mouridsen H, Andersen KW. Clinical trials. Semin Surg Oncol 1996; 12: 32–8.
13. Wohlfahrt J, Mouridsen H, Andersen PK, Melbye M. Reproductive risk factors for breast cancer by receptor status, histology, laterality and location. Int J Cancer 1999; 81: 49–55.