

0959-8049(95)00501-3

Original Paper

Quality of Prostate Cancer Data in the Cancer Registry of Norway

S. Harvei, S. Tretli and F. Langmark

Cancer Registry of Norway, Institute of Epidemiological Cancer Research, 0310 Oslo, Norway

Completeness of reporting and internal validity of the coding of prostate cancer in the Cancer Registry (CR) in Norway were examined. Data were matched and evaluated against diagnostic indices at eight selected hospitals in the country and against death certificates from Statistics Norway. Validity control was based on detailed re-analysis of an approximately 1% sample of the registered data during the period 1957–1986. The deficiency in reporting of prostate cancer was less than 1%. The grave deficiencies in hospital patient registers were considered to be of non-systematic nature and should, therefore, not impair the reliability of our investigation of incompleteness. The validity control revealed errors in 0.5% of the data elements, or, illustrated differently, 6% of the patient files had an error, of importance or not, in one of the data elements. One false positive registration was found among 298 controlled patient files (0.3%).

Key words: completeness, validity, prostate, registration, Norway

Eur J Cancer, Vol. 32A, No. 1, pp. 104–110, 1996

INTRODUCTION

POPULATION-BASED CANCER registries have an increasingly important role for cancer control programmes in society [1–4]. With the growing reliance on cancer registries databases in clinical and epidemiological cancer research, and to enable planning and evaluation of preventive and therapeutic services, it becomes increasingly necessary to assess the quality of the data. A Norwegian Public Health Act passed in 1983 has made each and every executor in the health services, individuals as well as institutions, responsible for evaluation and quality control. Therefore, in Norway, the Cancer Registry (CR) has an explicit legal obligation to assess the quality of its own registrations.

When studying the epidemiology of prostate cancer, one soon becomes aware of the fairly large variations between countries (registries) in both incidence and mortality. This indicates firstly a need to take a closer look at the quality of the registered data and secondly to secure standardised methods of diagnosis and registration.

The CR envisages a more intensive international epidemiological effort in the years to come in connection with prostate cancer—the second most frequent cancer in Norway. For this reason, the CR has decided to give priority to more thorough investigations of this site. We consider it our duty to provide scientists, political decision-makers and the public with high quality data, where weaknesses, possible errors and bias should

be known and presented to the reader. The objectives of this study were, therefore, to examine the completeness of the reporting of diagnosed cancer of the prostate in Norway and to assess the internal validity (reliability, reproducibility) of the coding at the CR.

Reporting of cancer has been compulsory in Norway since 1952 [5]. Reports are sent to the CR independently by hospital departments and histopathological laboratories, and the registered data are matched against incidence date and cause of death from the National Register of Deaths at Statistics Norway (the National Office of Statistics). This system of complementary acquisition of data from three different sources guarantees a high level of completeness. Since the late 1960s, pathologists have reported their results to the CR by means of photocopies of the standard forms for cytology, histology and autopsy which provide valuable additional clinical information. This information is also important in the efforts to obtain correct registration of clinical data as well as being very useful for internal checks of consistency. All coding is done centrally at the CR by trained coders.

An essential feature of the CR is the use of a national 11-digit personal identification number allocated to all Norwegian citizens from 1966. This unique identification number enables high quality book-keeping and record control and facilitates research. It is both effective in helping to prevent duplicate registration and is useful in follow-up procedures. This number makes a match (record linkage) between hospitals and the registry relatively easy.

The coding of diagnosis conforms with the *International*

Correspondence to S. Harvei.

Revised 24 Jul. 1995; accepted 4 Sep. 1995.

Classification of Diseases (ICD), 9th edition. Since 1970, morphology has been coded using the four-digit system of the *Systematised Nomenclature of Pathology* (SNOP) [6], and the old material from before 1970 has been recoded to SNOP codes as far as possible. Through the late 1970s and in the 1980s, the morphology code was adapted and restructured to the ICD-0 [7].

Essential elements in evaluating the quality of the data in the registry are the validity of the data, the reliability of coding and the completeness of case ascertainment. Validity of data depicts the accuracy of the diagnosis in the registry [8]. Validity also includes the accuracy of the other types of data on the cases registered compared with reported information. Validity of coding (reliability, reproducibility) signifies the internal consistency.

The CR itself established a system of editing and control when registration was started in 1952, and until 1960 the key elements of identification were name, date of birth, gender and place of residence. From 1982 and onwards the registration became online to the database.

The burden on staff in executing data quality control has lessened over time. The work was definitely most resource-intensive before the introduction of the identification number, when the names and date of birth did not initially always have enough discriminating ability in themselves to clearly identify each individual.

Development of tailor-made computer systems over the last 10–15 years has made it possible to refine and expand the quality control at the CR, not least to avoid duplication of recording, to control the pattern of reporting from the pathological laboratories, and to disclose internal inconsistencies in the coding. A lack of reporting, or a marked change in the number of reports, will thus be the subject of further queries.

MATERIALS AND METHODS

To study the completeness of the registration, CR data were matched against diagnostic indices at eight selected hospitals and against death certificates from Statistics Norway. The hospitals were selected as follows: in the county of Finnmark, an area of low incidence of prostate cancer, the study included the two ordinary county hospitals and the regional hospital in the neighbouring county; in Oslo, an area of high incidence of prostate cancer, the examination covered the two largest city hospitals, one medium-sized city hospital, and the national university clinic in Oslo; one hospital was also chosen from a county with an intermediate level of incidence. This was a hospital which included a specialised urological department for many years. Together, these three geographical areas contain one-fifth of Norway's total population. Three years, 1960, 1975 and 1981, selected at random from each decade of the study period, were chosen for verification of reporting.

If CR data did not completely match the hospital diagnostic indices, the case was checked against the hospital medical record. If the identification number was not used by the hospital, the checks were made manually using name, date of birth, gender and address. The process flow-chart in Figure 1 shows the routine followed in the different circumstances.

Whenever a histopathological report or a death certificate is received and is not matched by a report from the clinical department that took care of the patient, an inquiry is sent to the hospital. The type, time and number of inquiries mailed to clinical departments were noted for the years 1987–1990.

The analyses were based on the clinical and histopathological

information as originally provided from the laboratories of pathology.

Validity control of data was based on detailed re-analysis of an approximate 1% sample of the registered data on prostate cancer during the period 1957–1986. Three birth dates were chosen at random, and the records of registered individuals born in any year on these dates were subjected to control. This led to a sample of 301 cases out of a total of 35 656 (0.84%). The characteristics of the sample and of the total database information on prostate cancer are shown in Table 1. The original clinical and histopathological reports, death certificates, letters, and correspondence on the patients in the control sample were found and reviewed.

A comparison was made between the classification and coding of the data existing in the updated computer files in the registry and the original basic reports, and all errors were noted. The personal identification number was checked against the files of Statistics Norway. This control included a check of names and residence. For patients who died before the census of 1960, and therefore had no national identification number, the registered data were checked manually against Statistics Norway's lists of causes of deaths in order to control identity. The following 12 data elements were controlled for every case: personal identification (names/identification number), basis (method) of diagnosis, residence (county and municipality), reliability code of diagnosis, time of diagnosis (month/year), time of death, primary site of cancer, surgery, morphology, radiotherapy, stage of metastasis, other therapy (hormones, chemotherapy etc.).

RESULTS

For the period 1957–1986, only 1.7% (5/301) of the patients were registered on the basis of a death certificate alone. At least one clinical and one histopathological report were available for 72%, while 17.3% of the registrations were based on one clinical report only and 4.6% on one histopathological report only.

Completeness and time lag in reporting

A total of 778 cases of cancer of the prostate were reported to the CR by the selected hospitals as newly diagnosed cases in the specific years, while hospital lists showed that in these same years, 1112 patients were admitted to these hospitals for prostate cancer (Table 2), either being diagnosed in that particular year (incident cases) or readmitted for previously diagnosed cancer of the prostate (prevalent cases). 39 of the 1112 patients admitted to hospital with a prostate cancer diagnosis were not at first identified in the CR files, because of incorrect date of birth, misprinted name or wrong diagnosis (16 cases). 36 of the 39 patient records were found and controlled, but one of these cases had not been reported to the CR. Three medical records were not traced.

Of the 778 patients reported to the CR by the selected hospitals in the relevant years, 136 were not found on hospital lists or on diagnostic indices. These cases were checked manually against patient records. Fifteen patient records were not traced, but the remaining 121 cases were found with the correct diagnosis. There were quite a number of cases of misdiagnosis in the diagnostic indices of the hospitals, but further examination was considered to be outside the scope of this study.

Completeness improves with number of years following registration: after 1 year 95%; after 5 years 99%; and 10 years after registration the figure was even closer to 100%.

The study gave no indication of any difference in the quality of reporting between high incidence and low incidence regions

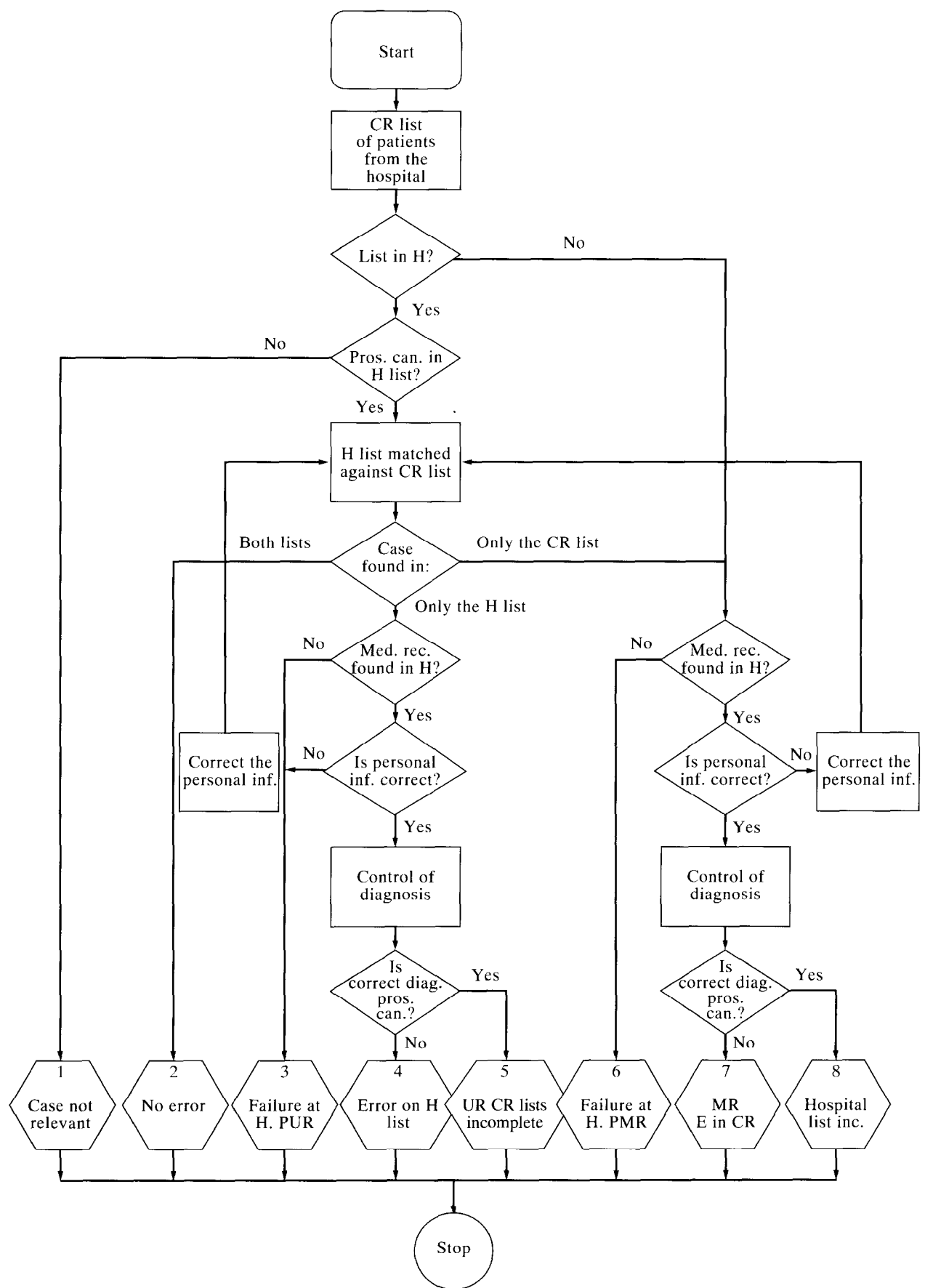


Figure 1. Process flow chart of control routines in this study. H, hospital; CR, Cancer Registry; Pros. can., prostate cancer; Med. rec., medical record; diag., diagnosis; info., information; PUR, possible unreported case; PMR, possible misreported case; MR, misreported case; E in CR, error in the Cancer Registry; UR, unreported case.

Table 1. Relationship between sample of 301 out of 35 656 prostate cancer cases in the database of the Cancer Registry of Norway

	Database		Sample	
	Number	Percent	Number	Percent
Year				
1957–1966	7803	21.9	59	19.6
1967–1976	11692	32.8	107	35.5
1977–1986	16161	45.3	135	44.9
Total	35656	100.0	301	100.0
Average age	73.6 years		73.3 years	
Metastatic status				
Localised	23439	65.8	203	67.4
Regional metastasis/ local infiltration	1521	4.3	8	2.7
Distant metastasis	8317	23.3	70	23.3
Metastasis NOS	331	0.9	1	0.3
Metastasis unknown	2048	5.7	19	6.3
Total	35656	100.0	301	100.0

NOS, not otherwise specified.

(Table 2). The large number of patients not found on hospital lists reflects the incomplete recording of diagnosis in hospitals, even in university hospitals. Naturally, this situation is a cause of concern with much wider perspectives than quality of cancer data and shows that, as yet, hospital lists or diagnostic indices are not reliable sources of information.

The percentage of the patient records in the CR which were found in the hospital discharge registers or in diagnostic files in the departments varied quite considerably. In the case of hospitals which had reported 10 or more cancer patients to the CR in 1981, the proportion of missing records varied between 5 and 20%. In 1960, only two-thirds of the CR patients were traced through the diagnostic registers, even at the large university hospitals.

Validity and reliability control

The checking of the 778 cases reported to the CR (Table 2) revealed no errors in the ICD codes (diagnosis) for the 763 cases that could be traced through the medical records. Two of the medical records that were not found referred to patients who had been reported to the CR after autopsy. It was frequently observed that when diagnosis was first made at autopsy, neither the hospital diagnostic indices nor the diagnosis on the medical record were updated. One patient on a hospital diagnostic list and not found originally in CR, was traced after a further search and the record in the hospital was revised to the correct diagnosis.

For the sample of 301 patients, the original report forms were found in the CR for 298. In 3 cases, no reports were found which could have provided the basis of the diagnosis, and it was impossible to check validity. The deficiency was attributed to internal office routines.

The matching against the National Register of Deaths did not bring to light any deficiency in the registration of prostate cancer in the regions concerned.

Table 3 gives the results of the validity control of 3411 elements of information on 298 patients and a detailed description of the types of error. In the case of 18 patients, one error was found in each of the patient records, that is, in 6% of the registered cases. Another measure of the validity is that 18 errors were found among the 3411 controlled data elements (0.5%).

There were no errors in the registration of personal identification: name, date of birth or personal identification number. The most serious error discovered was a case which was registered as cancer but was actually an adenoma. This was primarily a coding error, but secondarily a failure of the updating procedures. Only in 2 other cases did the system function inadequately, and most of the errors were coding failures. The majority of the errors were connected to coding of metastases and the basis of diagnosis, the main reason being the need to combine information from various reports when deciding the codes. Many of the errors can be considered of minor importance.

DISCUSSION

This study evaluates the completeness and validity of prostate cancer data in the CR, but does not assess the accuracy of the underlying medical record. It, therefore, provides no check on the quality and changes of the medical (diagnostic) services. The latter type of quality control, which could be termed external validity control of data [9, 10], is not discussed in this paper but is dealt with in a separate publication in this issue (pp. 111–117).

An important part of quality assurance and control is the preservation of confidentiality. The Data Inspectorate has set the rules and limitations for the handling of person-identified medical data. All the activities necessary to carry out this study were performed in a way that specifically ensured confidential handling of data.

The greater the proportion of histologically confirmed cases and the more multiple reports we receive on each case, the greater is our confidence in the quality of our data. The rate of histological and cytological verification is now higher than 95% and is among the highest rates reported to the *Cancer Incidence in Five Continents* 1992 [11].

A low percentage of reporting based solely on death certificates has also been regarded as a sign of high quality. For prostate cancer, the present figure is still 1.4%, reflecting that this is a form of cancer especially associated with old age. However, even this figure is relatively small compared with that reported from other registries [11].

The most comprehensive test of completeness of the reporting of all types of cancer to the CR stems from a national patient census carried out on 1 October 1970 in all health institutions in Norway. Diagnosis was stated for all types of cancer in all patients in hospitals, nursing homes, convalescence homes and hospices. A total of 2085 patients with a cancer diagnosis were identified and matched against the CR file in 1981 [12]. Only 19 patients were identified with a cancer diagnosis that should have been reported to the CR, which is 0.9% (95% confidence interval 0.6–1.4%). Therefore, our most realistic estimate is currently 1% incompleteness in the reporting of all cancers to the CR.

The incompleteness in reporting of prostate cancer as indicated from the numbers in the present study, is one out of 1112 records, which is astonishingly low. Even if we assume that the three medical records with prostate cancer on the diagnostic index, but not found in the CR, were correctly diagnosed, the sum would still be only four missing reports for a total of 1112 patients, a proportion of error of 0.4% (95% confidence interval 0.1–0.9%). This coincides with our belief that the reporting of prostate cancer is better than for many other cancers, owing to the small range of histological types and a high degree of histological/cytological verification. We have found no parallel projects that directly assess the completeness of prostate cancer data in cancer registries. An evaluation of the registration of all

Table 2. Matching of Cancer Registry (CR) files against hospital lists and vice versa

Hospital	Incidence	Year	Patient on hospital lists					Cancer Registry lists				
			Number of cases		Medical records			Number of cases		Medical records		
			On the hospital lists	Only on the hospital lists	Not found	Found		Reported from the hospital	Only on the CR lists	Not found	Found	
						Diagnosis not correct	Diagnosis correct				Diagnosis not correct	Diagnosis correct
A (univ.)	High	1960	58	3	0	2	1	30	9	0	0	9
		1975	87	6	1	0	5	60	12	1	0	11
		1981	135	3	0	0	3	63	4	0	0	4
B (univ.)	High	1960	32	8	0	3	5	34	8	2	0	6
		1975	32	2	0	1	1‡	31	2	0	0	2
		1981	49	6	0	6	0	28	5	1	0	4
C (univ.)	High	1960	22	0	0	0	0	31	11	2	0	9
		1975	150	2	2	0	0	96	6	0	0	6
		1981	164	6	0	2	4	114	5	0	0	5
D (local)	High	1960	—	—	—	—	—	2	2	1	0	1
		1975	—	—	—	—	—	19	19	1	0	18
		1981	14	0	0	0	0	10	2	0	0	2
E (reg.)	Intermed.	1960	40	1	0	1	0	24	3	0	0	3
		1975	95	1	0	0	1	50	2	0	0	2
		1981	129	0	0	0	0	88	14	3	0	11
F (univ./reg.)	Low	1961*	15	1	0	1	0	16	3	2	0	1
		1975	—	—	—	—	—	19	19	0	0	19
		1981	65	0	0	0	0	40	1	1	0	0
G (local)	Low	1960	2	0	0	0	0	2	2	0	0	2
		1975	3	0	0	0	0	6	2	1	0	1
		1981	13	0	0	0	0	10	2	0	0	2
H (local)	Low	1960	1	0	0	0	0	1	0	0	0	0
		1975	1	0	0	0	0	1	1	0	0	1
		1981	5	0	0	0	0	3	2	0	0	2
Total			1112	39	3	16	20	778	136	15	0	121
Correspondence with Figure 1.					3	4	5‡			6	7	8

*1960 not available. †All these cases except (‡) matched with CR lists after the patient was identified correctly. ‡This is the only case not reported to the CR. —, no patient lists. univ., university; reg., regional; Intermed, intermediate.

cases of cancer in the Danish Cancer Registry in 1977 [13] was undertaken by matching the registry data against the data of the National Patient Register. The deficiency of reporting of male genital cancer was 2%. A Swedish investigation of under-reporting to the Stockholm Cancer Registry in 1978 [14], carried out by matching against the regional in-patient diagnosis register, gave the same result as for Denmark: approximately 2% registration deficit for prostate cancer. Sweden had not previously registered cancer on the basis of death certificates, and an examination of the completeness of the Swedish National Cancer Registry based on non-notified cancer cases recorded on death certificates in 1978 [15] gave a deficit of 4.6% of prostate cancer cases in the CR. Our figures of incompleteness are basically lower than these figures and lower than those presented by many other registries [16–18]. A comparison with cancer registries outside the Nordic countries is difficult, mainly owing to differences in coverage, in the registration systems themselves and in the structure of the health services.

In Norway, there is neither a national diagnostic register of discharged hospital patients as in Denmark, nor any complete diagnostic registers—computerised or not—at the hospitals for

the relevant years. An important question is whether the demonstrated grave deficiencies in hospital diagnostic indices impairs the reliability of our investigation of completeness. However, we regard the incompleteness of the hospital diagnostic registers, or the absence of such registers, to be of an unsystematic nature. Therefore, we also consider the number of cancer patients in these eight hospitals to be sufficiently representative as regards the reporting of prostate cancer to the CR.

Validity of data—internal data quality assessment

Since the CR is a national population-based registry, no selection bias has to be searched for. Relevant validity control should be directed towards information bias—the checking of systematic and random errors [19]. In this connection, it is considered beneficial that the average number of reports on each patient is more than three. In the checking of consistency, comparisons are made between items on different records for the same patient, which secures a high level of validity of the data.

One of the 298 searched records was erroneously registered as cancer, equivalent to 0.3% false positive cases. This rate, if

Table 3. Results of validity control of Cancer Registry data on cancer of the prostate

Information elements (n = 3411)*	No. of errors	Type of errors	Comments
Date of birth/ID number	0	—	—
Site of cancer	0	—	—
Residency	2	Coding	Two records with wrong code of local municipality, one with correct and one with county code wrong. The regional incidence would not have been affected by these errors
Time of diagnosis	1	System	Date of death erroneously recorded by system as time of diagnosis.
Extended malignancy code	1	Coding	Erroneously coded as cancer
Histology	2	One coding/one combined coding/system	One record coded as cancer, correct: 'adenoma' One record coded 'malignant tumour', correct: 'carcinoma'
Metastasis	6	Coding	One record coded 'unknown metastatic status', correct: 'no metastasis' One record coded 'no metastasis', correct: 'not given' One record coded 'no metastasis', correct: 'local infiltration' One record coded 'metastasis unknown', correct: 'distant metastasis' One record coded 'no metastasis', correct: 'distant metastasis' One record coded 'distant metastasis', correct: 'unknown metastatic status'
Basis of diagnosis	4	Three coding, one system	Coded 'clinical examination', correct: 'endoscopic examination' Two records coded 'histological examination', correct: 'autopsy' One record coded 'biopsy', correct: 'cytology'
Treatment	2	Coding	One record coded 'transvesical operation', correct: 'palliative operation not directed towards primary tumour' One record coded 'transurethral prostatic resection', correct: 'biopsy only'
Total	18		

*Treatment items were not properly registered 1981–1985, and therefore not subject to validity control in this period (for $n = 55$ individuals). 243 cases (out of 298) were therefore controlled for 12 information elements each, while the 55 cases could be controlled for only 9 items (12–3), altogether 3411 elements $\{(243 \times 12) + (55 \times 9)\}$. ID, identification.

applied to the number of newly diagnosed cases of prostate cancer in 1991, would imply that 6 cases of the 1996 were falsely registered as cancer. This would have no significant influence on the calculation of either incidence or mortality statistics.

Most of the other errors occurred in the coding procedures and were impossible to discover by the logical controls performed by the computer. Only one of the 18 errors in codes was a combined error of coding and system failure. Approximately half the errors in the coding of 'metastasis' and 'basis of diagnosis' were unimportant. We are, therefore, left with a practical consideration of an error rate of approximately 1% of all data elements. It is difficult to envisage any situation in epidemiological analysis where random errors of this size have a significant influence on the conclusions. The Finnish Cancer Registry examined the error rate in the registration of colorectal cancer in 1975, and found that 5.9% of the cases were erroneous [16]. This is in agreement with our findings. An examination of the accuracy of cancer registration in 1988 was carried out at Tayside Cancer Registry in Scotland [20] by comparing 200 consecutive registrations of any cancer type with histopathology reports. Approximately 6% serious errors, 3% moderate errors and 17.5% minor errors in the ICD coding were found, when the histopathology report was taken as the "gold standard". It seems to reflect the discrepancies between the cancer registry files and the histopathology report, rather than give a true picture of the error rate in the cancer registry.

1. World Health Organization (WHO), Regional Office for Europe. The role of epidemiology in cancer control. ICP/CAN 015 0369H. Report on a WHO Meeting, Lyon, 2–4 November 1982.
2. Knowelden J, Mork T, Philips AJ, eds. *The Registry in Cancer Control*. UICC Technical Report Series, Vol V, Geneva, 1970.
3. Klawansky S, Burdick E, Adams M, Bollini P, Orza M, Falotico-Taylor J. Use of the SEER Cancer Registry for technology assessment. *Int J Technol Assess Health Care* 1991, 7, 134–142.
4. Muir CS, Démaret E, Boyle P. The cancer registry in cancer control: an overview. In Parkin DM, Wagner G, Muir C, eds. *The Role of the Registry in Cancer Control*. IARC Scientific Publications, Lyon, IARC, 1985, No. 66.
5. The Cancer Registry of Norway (eds Pedersen E, Magnus K). *Cancer Registration in Norway. The Incidence of Cancer in Norway 1953–1954*. The Norwegian Cancer Society, Monograph No. 1, Oslo, 1959.
6. *Systematized Nomenclature of Pathology (SNOP)*. College of American Pathologists, Chicago, 1965.
7. WHO, ICD-0, *International Classification of Diseases for Oncology*. World Health Organization, Geneva, 1976.
8. Kelsey JL, Douglas Thompson W, Evans AS. *Methods in Observational Epidemiology. Monographs in Epidemiology and Biostatistics*. Oxford University Press, 1986, Vol. 10.
9. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research—Principles and Quantitative Methods*. Van Nostrand Reinhold Company 1982, 183–193.
10. Rothman KJ. *Modern Epidemiology*. Little, Brown and Company 1986, 82–97.
11. Parkin DM, Muir CS, Whelan SL, Gao Y-T, Ferlay J, Powell J, eds. *Cancer Incidence in Five Continents*. IARC Scientific Publications, Vol. VI, No. 120, Lyon, IARC, 1992.

12. The Cancer Registry of Norway. Evaluation of completeness of reporting. In *Annual Report of the Cancer Registry 1982*, 19–20 (in Norwegian).
13. Österlind A, Jensen OM. Evaluation of registration of cancer cases in Denmark 1977. Preliminary evaluation of registration of cases by the cancer registry and the National Patient Register. *Ugeskr Laeger* 1985, **147**, 2483–2488 (in Danish).
14. Mattsson B, Rutqvist LE, Wallgren A. Undernotification of diagnosed cancer cases to the Stockholm Cancer Registry. *Int J Epidemiol* 1985, **14**, 64–69.
15. Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. *Acta Radiol Oncol* 1984, **23**, 305–313.
16. Kyllonen LE, Teppo L, Lehtonen M. Completeness and accuracy of registration of colorectal cancer in Finland. *Ann Chir Gynaecol* 1987, **76**, 185–190.
17. Mukherjee AK, Leck I, Langley FA, Ashcroft C. The completeness and accuracy of health authority and cancer registry records according to a study of ovarian neoplasms. *Public Health* 1991, **105**, 69–78.
18. Benn RT, Leck I, Nwene UP. Estimation of completeness of cancer registration. *Int J Epidemiol* 1982, **11**, 362–367.
19. Rothman KJ. *Modern Epidemiology*. Little, Brown and Company 1986, 77–97.
20. Lapham R, Waugh NR. An audit of the quality of cancer registration data. *Br J Cancer* 1992, **66**, 552–554.