available at www.sciencedirect.com

ScienceDirect

journal homepage: www.ejconline.com

Data quality at the Cancer Registry of Norway: An overview of comparability, completeness, validity and timeliness

Inger Kristin Larsen[a,*], *Milada Småstuen*[a], *Tom Børge Johannesen*[a], *Frøydis Langmark*[a], *Donald Maxwell Parkin*[c,d], *Freddie Bray*[a,b], *Bjørn Møller*[a]

[a]*Department of Clinical and Registry-based Research, Cancer Registry of Norway, Institute of Population-based Cancer Research, Fridjof Nansens vei 19, Postbox 5313, Majorstuen, N-0304 Oslo, Norway*
[b]*Department of Biostatistics, Institute of Basic Medical Sciences, University of Oslo, Norway*
[c]*Clinical Trials Service Unit and Epidemiological Studies Unit, University of Oxford, Oxford, UK*
[d]*Cancer Research UK, Centre for Epidemiology, Mathematics and Statistics, London, UK*

ARTICLE INFO

Article history:
Received 1 September 2008
Received in revised form 24 October 2008
Accepted 28 October 2008
Available online 16 December 2008

Keywords:
Data quality
Cancer Registry
Neoplasms
Comparability
Completeness
Validity
Accuracy
Timeliness

ABSTRACT

Aim: To provide a comprehensive evaluation of the quality of the data collected on both solid and non-solid tumours at the Cancer Registry of Norway (CRN).
Methods: Established quantitative and semi-quantitative methods were used to assess comparability, completeness, accuracy and timeliness of data for the period 1953–2005, with special attention to the registration period 2001–2005.
Results: The CRN coding and classification system by and large follows international standards, with some further subdivisions of morphology groupings performed in-house. The overall completeness was estimated at 98.8% for the registration period 2001–2005. There remains a variable degree of under-reporting particularly for haematological malignancies (C90–95) and tumours of the central nervous system (C70–72). For the same period, 93.8% of the cases were morphologically verified (site-specific range: 60.0–99.8%). The under-reporting in 2005 due to timely publication is estimated at 2.2% overall, based on the number of cases received at the registry during the following year.
Conclusion: This review suggests the routines in place at the CRN yields comparable data that can be considered reasonably accurate, close-to-complete and timely, thereby justifying our policy of the reporting of annual incidence one year after the year of diagnosis.

1. Introduction

The reporting of neoplasms (and certain precancerous lesions) to the Cancer Registry of Norway (CRN) has been compulsory following a directive from the Ministry of Health and Social Affairs in 1951. A Health Registry Act came into force in 2002 that included statutory regulations for the CRN,[1] strengthening the legal obligation to report new precancerous and cancerous cases, and requiring all hospitals, laboratories and general practitioners in Norway to report these cases. This legislation, the efficient reporting from multiple sources and the trace-back routines have been built up at the Registry over the last decades.

Several studies at the regional[2] and site-specific level[3–5] have reported an overall estimated of completeness at the CRN at greater than 95%. An overview of these studies is given in the Appendix. An evaluation of the data quality at CRN was recently included as a special issue of the Registry's annual report *Cancer in Norway*.[6] Selected results from that evaluation are elaborated and discussed in this paper.

* *Corresponding author:* Tel.: +47 22 45 13 00/+47 93 42 43 74; fax: + 47 22 45 13 70.
E-mail address: inger.kristin.larsen@kreftregisteret.no (I.K. Larsen).
0959-8049/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.ejca.2008.10.037

The overall aim of this study is to provide a comprehensive evaluation of the data quality of the data at the CRN for all cancer types, applying the principles and techniques of evaluation described in the accompanying reports on data quality methods in this issue.[7,8]

2. Material

2.1. Sources of information in the Cancer Registry of Norway

Hospitals, pathological laboratories, general practitioners and the National Statistics Office (*Statistics Norway*) provide the key information that enables the CRN to collect, code and store data on cancer patients in Norway. Clinical notifications are reported on structured templates with one form for solid cancers and another for non-solid tumours. Information is sought on localisation, the extent of the disease and treatment on these forms. In addition, some site-specific forms with additional clinical information exist for malignant lymphoma, chronic lymphatic leukaemia, and cancers of the breast, ovary, prostate and colorectum. Pathology notifications from all the laboratories are copies of the laboratory workers' own reports, which are routinely sent to the CRN. The pathological notifications provide either histological, haematological, cytological or autopsy information. *Statistics Norway* provides information on the cause of death notified on death certificates (DCs), and updates monthly the vital status (alive, dead or emigrated) of all registered persons in the CRN.

The CRN has also received data files with discharge diagnosis from 1998 on all patients treated for malignant and pre-malignant conditions in every Norwegian hospital and outpatient clinic. The discharge diagnoses are used as basis for sending out reminders for clinical notification.

Three times a year, reminders are sent to all hospitals and physicians that have failed to initially report new cases within 2 months of diagnosis or have given insufficient information for registration. Elderly patients residing in Nursing homes are generally diagnosed in the hospital. Reminders are, however, sent to the nursing home where the pathological notification states them as the clinical contact institution, or if a cancer diagnosis is mentioned on the death certificate and no additional information is available from hospitals or pathological laboratories.

In addition to the discharge diagnoses, laboratory reports and the mentioning of cancer on the death certificate provide the basis for these reminders.

As with the other Nordic countries, patients are identified through a unique 11-digit personal identification number. This number is assigned to all newborns and people residing in Norway, and through the consequent effectiveness of record linkage, provides a reliable means of tracking patients, and keeping the potential duplication of registered cases down to a minimum.

2.2. Data material in this study

Data for the period 1953–2006 were extracted from the CRN. For the evaluation of comparability, the whole time span was used, whereas selected recent years were used in the application of methods for the evaluation of validity, completeness and timeliness. The specific years utilised are given in the methods section. The dataset for the period 2001–2005 comprised 119,049 incident cases diagnosed in 114,381 persons. Analyses are presented at the two-digit ICD-10 level following a translation from ICD-7, with the codes C39, C76 and C80 defined as 'other or unspecified' cancers. Population data, stratified by year, sex and age were provided by Statistics Norway.

3. Methods

A brief overview of the subset of techniques covered by Bray and Parkin and Parkin and Bray[7,8] that have been applied to the data from the CRN is set out below, together with a specification of the relevant time periods used for each method.

3.1. Comparability

Comparability is the extent to which coding and classification procedures at the CRN, together with the definitions of recording and recording specific data items adhere to agreed international guidelines. The topics covered here are international standards for classification and coding of neoplasms (1953–2006), as well as the definitions of incidence, incidence date and multiple primaries.

3.2. Completeness

Completeness is defined as the extent to which all diagnosed neoplasms in Norway are included in the registry database. Three semi-quantitative methods are included here, namely the historic data method (stability of data over time (1953–2006) and age-specific incidence rates of childhood cancer (2001–2005) with the corresponding reference intervals based on deciles for childhood cancer published in CI5 Volume VIII[9] (Table 1 of Parkin and Bray[8])), the mortality: incidence ratio (the ratio 2000–2004 compared with one minus five year relative survival for cases diagnosed in 1996–2000) and the number of sources/notifications per case (for the registration period 2001–2005). Among the quantitative methods outlined by Parkin and Bray,[8] we have applied the capture–recapture (cases diagnosed in 2001–2005) and the 'flow' method (cases diagnosed in 1999 and followed up to the end of 2004).

3.3. Validity (accuracy)

Validity is defined as the proportion of cases in the CRN with a given characteristic which truly have this attribute. Among the validity measures described by Bray and Parkin,[7] the proportions with histological verification (for registrations 2001–2005), obtained through death certificate sources only (2001–2005) and missing information (the proportion of cases registered with primary site unknown (PSU), by age groups, and the proportion with stage unknown, by site, in the period 2001–2005) are presented, and compared with selected European registries for patients diagnosed in 1998–2002, as part of the compilation of *Cancer Incidence in Five Continents* (CI5), Volume IX.[10]

3.4. *Timeliness*

Timeliness is evaluated in terms of the time from diagnosis to registration, and the time from registration to the reporting of incidence via the annual report, *Cancer in Norway*.

4. Results

4.1. *Comparability*

Incident cases in Norway comprise all malignant and *in situ* neoplasms, and the incidence reported in Cancer in Norway (CiN) includes all cases with the 5th-digit behaviour code 3 according to ICD-O-3, for haematological malignancies, and ICD-O-2, for other tumours. Cases with 5th-digit behaviour code 1 are also included for tumours of the central nervous system.

Fig. 1 gives an overview of the international standards that have been followed[11–16] for the classification and coding of neoplasms from the start of cancer registration in 1953 up to the latest year available at the time of publication, 2006. The topography codes have been converted to ICD-10[17] for reporting purposes. Some morphology codes have been developed due to the need for further subdivision of ICD-O-2 and ICD-O-3 on the basis of additional information on morphology, cytochemistry and immunophenotype, especially within the haematological malignancies. These subdivisions are listed in ICD-O-3 as separate terms, but with the same morphology code.

Table 1 – Age-specific incidence rates per 100,000 for childhood cancer by gender, Norway, 2001–2005.

Age	Boys	Reference	Girls	Reference
0–4	23.5	(<12.3 – >24.7)	19.5	(<9.7 – >21.4)
5–9	12.3	(<8.5 – >15.6)	14.0	(<6.9 – >12.0)
10–14	15.3	(<8.5 – >15.0)	11.1	(<6.8 – >13.6)

The in-house rules for the registration of incidence date (for the reporting of new cases or calculating survival) depart from the European Network of Cancer Registries (ENCR) recommendations,[17] as the Registry always registers the earliest incidence dates reported on the sources of notification, whereas the ENCR rules are based upon a hierarchy of possible sources. For the period 2001–2005, 19.7% of the cases had a different date on applying the ENCR rules. For these cases, the median difference between the ENCR-defined date and the in-house incidence date was 10 days.

For reporting and comparability with other registries it is, however, possible, when needed, to select and give priority to the date when the specimen was taken.

The recording of multiple primary tumours in the main follows the recommendations given by ENCR.[18] The recognition of two or more primary cancers does not depend on time, and the groups of topography codes considered as single sites (from ICD-O-2 and ICD-O-3) are followed, with systemic and multicentric cancers counted only once. The CRN has, however, used a more detailed grouping of specific histologies considered to be different for the purposes of defining multiple tumours. If a new tumour has a different morphological code (e.g. the first four of the six digits denote a different cell type), it is considered by an in-house physician for possible classification as a new cancer. Also, there is a further consideration of whether two tumour diagnoses with identical morphological codes for the same site more than 4 months apart constitute a new tumour or a recurrence. For the purposes of incidence reporting, the yearly publication from the CRN includes the first primary tumour within the same three character categories of the topography code in each patient.

4.2. *Completeness*

The assessment of historic incidence trends, by sex, for some selected cancers is shown in Fig. 2. While the gradient varies with cancer, the annual trends do not appear to fluctuate in any systematic pattern. The age-specific incidence rates for

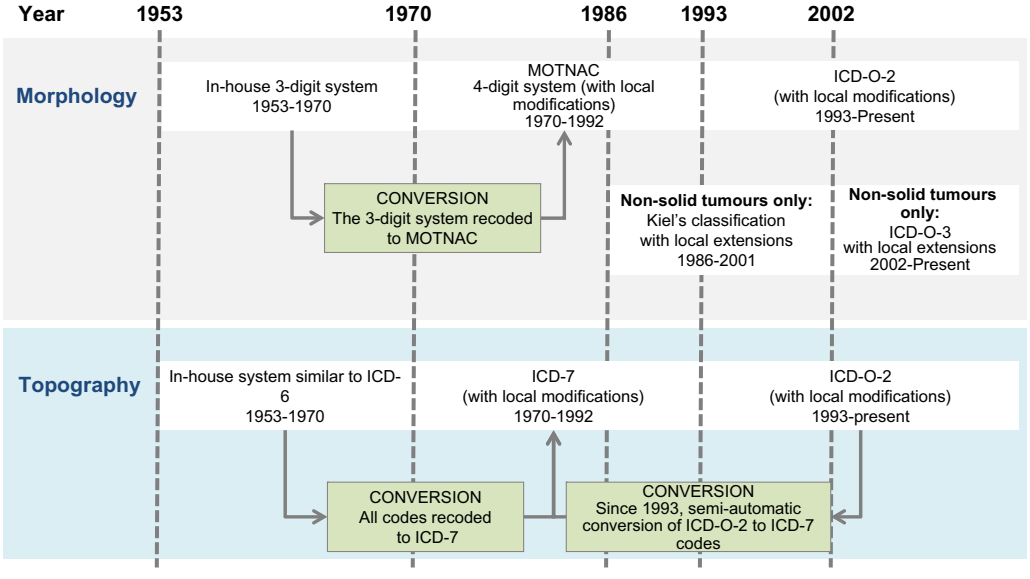

Fig. 1 – The standards of classification and coding of neoplasms followed, Norway 1953–2007.

childhood cancer in the period 2001–2005 are shown by sex in Table 1. The values for girls aged 5–9 years were outside the upper limit of the reference interval.

In general, there was a strong coherence between the M:I ratio and 1 – overall survival, e.g. cancers with poor survival as pancreas, liver, oesophagus and lung cancers had M:I ratios close to 1 (Fig. 3). Discrepancies between the two variables were mostly in the direction of M:I being higher the 1-survival. Fig. 4 shows a comparison of the M:I ratios for all cancer sites for diagnoses in 2000–2004 in Norway and Finland. There appears to be reasonable concordance in the M:I ratios between the two countries, although a statistical test for variations according to neoplasm flagged cancers of the gallbladder, pleura, penis, bladder, eye, lip, colon and prostate, as well as other leukaemia, multiple myeloma and melanoma of skin, as significantly different (data not shown).

The average number of notifications per case, for all sites combined, was 3.2 (Table 2). Breast and cervical cancers were among sites with the highest rate of notifications per case, whereas cancers of the central nervous system, kidney or

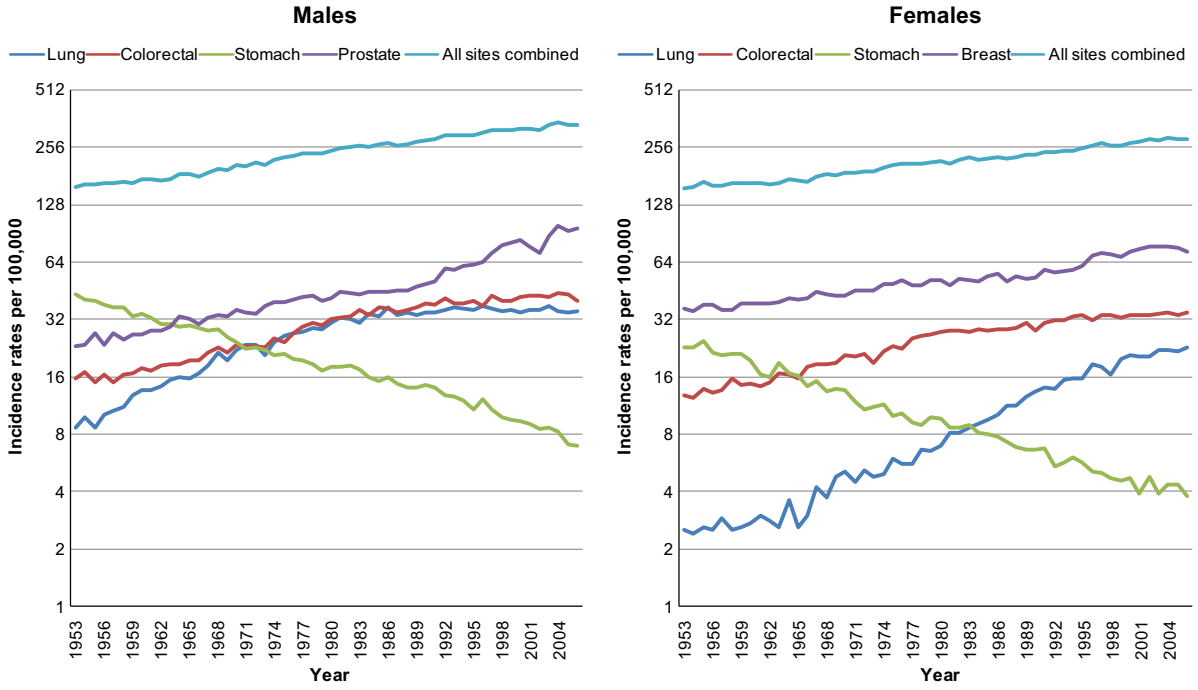

Fig. 2 – Annual trends in age-standardised (world) incidence rates for all sites combined, and for selected sites, 1953–2005, Norway.

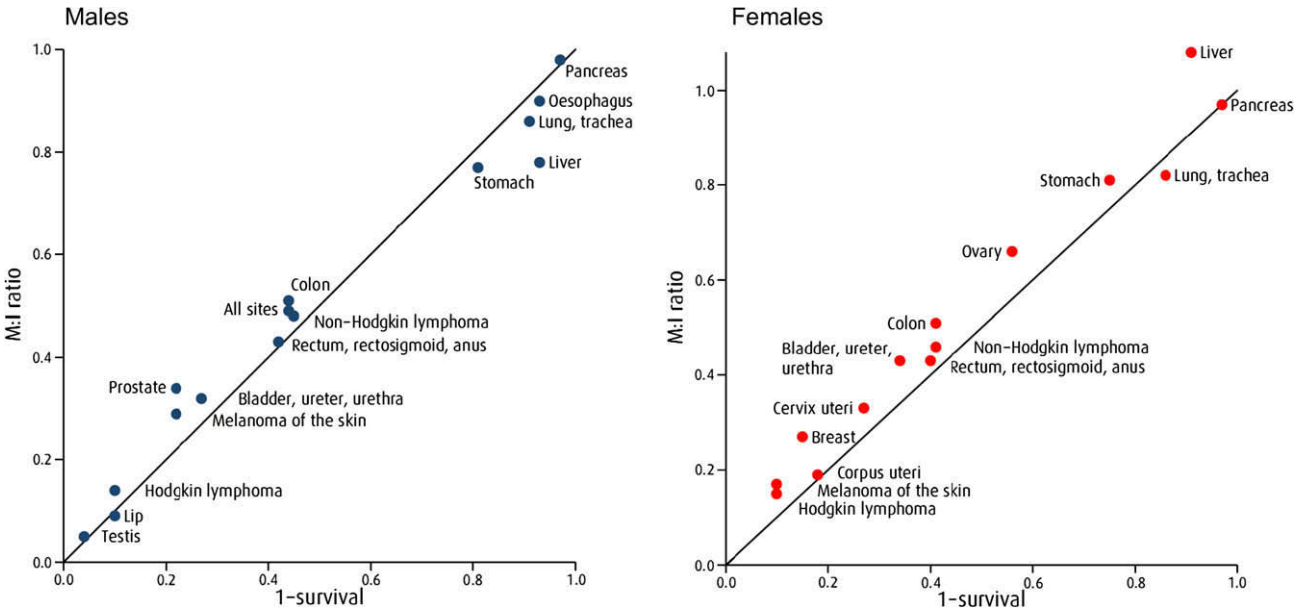

Fig. 3 – Mortality:incidence ratios (2000–2004) versus 1-survival (based on diagnoses in 1996–2000), Norway.

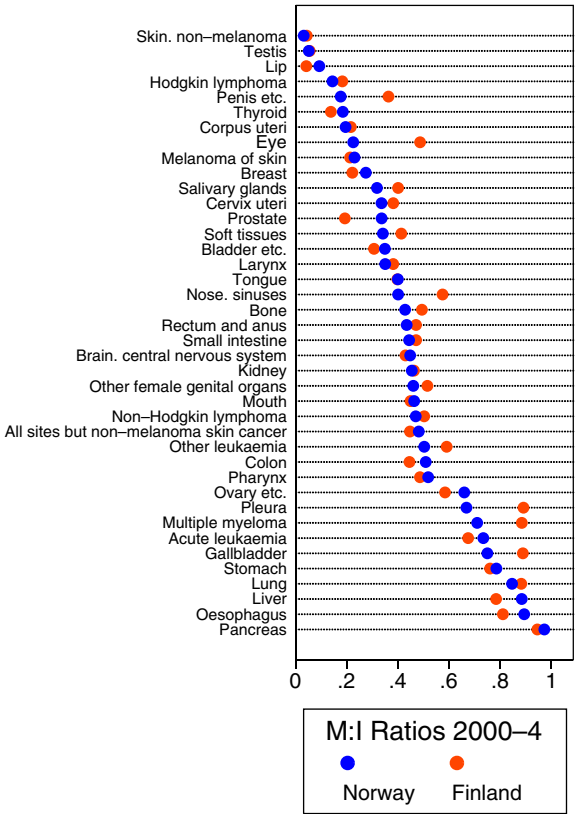

Fig. 4 – Comparison of mortality:incidence ratios by cancer site, Norway versus Finland, 2000–2004.

multiple myeloma were among sites registered with a low number of notifications per case.

Estimates of completeness, using the capture–recapture method, are given in Table 3. For the period 2001–2005, and for all sites combined, 93% of the cases notified by clinicians (n = 103,644) were also registered with information from at least one other source. For pathological notifications (n = 99,552) or cases notified via a DC (n = 17,786), notifications from one of the other two sources were present for 91.4% and 93.1% of the cases, respectively. In the same period, a total of 11,671 (9.8%) cases were registered without a clinical notification (ranging from 5.4% in 2001 to 19.3% in 2005). For most cancers, the dependencies were (weakly) positive between notifications by clinicians and DCs and also between notifications from clinicians and pathologists. The dependencies between DCs and pathology notifications were mostly negative.

The overall completeness for the period 2001–2005, estimated by the capture/recapture method, was 98.8%. The lowest completeness was estimated for pancreas (95.7%), multiple myeloma (95.5%), leukaemia (94.6%) and central nervous system (93.8%). The overall completeness 5 years from the year of diagnosis in 1999, as estimated by the flow method, was 97.8% (Fig. 5). This Figure shows the completeness, for all sites combined, estimated with the flow method for cancer cases diagnosed in 1999 and who were followed up until 31st December 2004. Overall completeness for these cases was low for the first and second year of registration, but increased to well over 95% during the third year.

4.3. *Validity (accuracy)*

A total of 93.8% of the cancer cases registered in the period 2001–2005 were morphologically verified (Table 2). Fig. 6 shows that the proportion of morphologically verified cases in Norway compared to other European registries.

The proportion of DCO cases 2001–2005 was 0.9%, and there were some variations according to cancer site (Table 2). Three-quarters of the cancer sites listed had %DCO less than 1%; the highest proportions included pancreatic (3.6%) and liver cancer (3.3%). Fig. 7 shows the comparison of %DCO, for all sites combined, with other European registries. The DCO proportions in Norway rank amongst the lowest in Europe, irrespective of sex.

The overall proportion of cases registered with PSU was 2.2% in the period 2001–2005. Sub-analyses showed that the proportion of PSU was very low in the age-group 0–44 (0.4%). It increased to 1.3% and 1.5% for the age-groups 45–64 and 65–74, respectively, and then peaked at 3.6% for the oldest age-group (⩾75 years). Comparisons of %PSU in Norway to other European registries are shown in Fig. 8.

The proportion with stage unknown for diagnoses in 2001–2005 varies considerably by cancer site, ranging from 42% with missing stage information for prostate cancer to 4% for cervical cancer to 0.5% for female breast cancer (data not shown).

4.4. *Timeliness*

The median time from the date of diagnosis to a registration of a new case has been reduced from over 525 days in 2001 to 261 days in 2005. Table 4 shows the number of cancer cases diagnosed in the year 2005 as enumerated on 29th November 2006 (time of first publication), and 19th November 2007 (about one year later). The difference of 2.2% overall gives an estimate of the under-reporting in 2005 due to early publication.

5. Discussion

The prerequisites for population-based high quality cancer incidence data are favourable in Norway with mandatory reporting, unique personal identification numbers and more than 50 years of experience in cancer registration. The present evaluation of the quality of data supports the notion that the Registry has a high degree of comparability, completeness, accuracy and timeliness. Some heterogeneity was, however, observed for some quality indicators for specific cancer sites. Cancers of the breast, cervical, colorectal and prostate cancer were found to have very favourable values for all aspects assessed, whereas leukaemias generally and cancers of the central nervous system and pancreas did not meet the highest standards.

The overall completeness estimates from the present study indicate that recent registration data from Norway are among the most complete among the European Registries. In the comparison of quality indicators with other European registries, the CRN had results equivalent to the other Nordic

Table 2 – Number of cases, average number of notifications per case, percentage morphologically (%MV), and percentage obtained from death certificate (%DCO), Norway, 2001–2005.

ICD-10	Site	Cases	Notificat. per case	MV%	%DCO
C00–96	All sites	119049	3.2	93.8	0.9
C00	Lip	373	3.5	99.7	0.0
C01–02	Tongue	389	4.7	98.7	0.0
C03–06	Mouth, other	404	4.7	99.8	0.0
C07–08	Salivary glands	199	4.1	98.5	0.0
C09–14	Pharynx	588	4.5	99.0	0.2
C15	Oesophagus	934	3.3	97.1	1.4
C16	Stomach	2833	3.2	97.7	0.6
C17	Small intestine	510	3.0	98.1	0.0
C18	Colon	10954	2.8	96.6	0.7
C19–21	Rectum, rectosigmoid, anus	6051	3.3	98.4	0.4
C22	Liver	627	2.5	85.0	3.3
C23–24	Gallbladder, bile ducts	681	2.7	85.3	1.5
C25	Pancreas	3161	2.3	69.8	3.6
C26	Other digestive organs	438	2.4	65.2	11.4
C30–31	Nose, sinuses	201	4.7	99.5	0.0
C32	Larynx, epiglottis	609	4.1	99.0	0.2
C33–34	Lung, trachea	11235	3.2	90.4	1.5
C38	Mediastinum, pleura	100	3.2	72.4	4.0
C40–41	Bone	216	4.8	97.2	0.5
C43	Melanoma of the skin	5255	3.0	99.5	0.2
C44	Skin, non-melanoma	6040	2.3	99.8	0.0
C45	Mesothelioma	360	3.8	97.5	0.0
C46	Kaposi's sarcoma	46	2.4	93.5	0.0
C47	Autonomic nervous system	47	4.6	97.9	0.0
C48–49	Soft tissues	625	4.6	98.1	0.0
C50	Breast	13701	4.3	99.3	0.2
C53	Cervix uteri	1469	5.9	99.8	0.1
C54	Corpus uteri	3160	4.7	99.2	0.2
C55	Uterus, other	46	3.0	82.6	4.3
C56	Ovary	2270	4.8	95.3	0.8
C51–52,C57	Other female genital	662	5.0	96.4	2.0
C58	Placenta	20	2.9	60.0	0.0
C61	Prostate	16559	2.5	97.1	0.9
C62	Testis	1280	3.0	99.5	0.2
C60, C63	Other male genital	220	4.2	98.2	0.9
C64	Kidney excluding renal pelvis	2761	2.3	85.5	1.1
C65	Renal pelvis	340	2.8	98.0	0.0
C66–68	Bladder, ureter, urethra	6270	3.2	98.2	0.6
C69	Eye	308	2.4	71.2	0.3
C70–72	Central nervous system	4572	2.1	63.2	0.9
C73	Thyroid gland	1000	3.6	99.1	0.2
C37, C74–75	Other endocrine glands	118	3.7	81.4	2.5
C39, C76, C80	Other or unspecified	2614	2.4	64.6	5.9
C81	Hodgkin lymphoma	561	3.7	99.6	0.2
C82–85, C96	Non-Hodgkin lymphoma	3663	4.3	98.9	0.3
C88	Malignant immunoproliferative diseases	221	2.1	97.3	2.3
C90	Multiple myeloma	1680	2.1	97.2	1.8
C91–95	Leukaemia	2678	2.6	97.7	2.0

registries, all of whom have reported completeness in the range of 96–100%.[19–24]

The high degree of completeness is supported by earlier Norwegian studies investigating cancers of the prostate, head and neck and ovaries.[3–5] In these studies, more than 99% of all patients in hospital files were also registered in the cancer registry. The study by Lund from 1981, involving 12% of the total Norwegian population in 1976, indicated that the overall completeness in the CRN was 97.6%, with a very high completeness for most solid tumours.[2] Regarding the quantitative methods, the completeness estimates from the *flow method* were lower than the estimates from the *capture–recapture* method. A limitation of the former method is that it relies on the follow-up of vital status several years after the date of diagnosis. Thus, completeness was based on cases diagnosed in the year 1999, whereas the capture–recapture method estimated completeness for cases diagnosed in 2001–2005. Both methods are based on the assumptions that are not directly verifiable. The capture–recapture method has been criticised because it assumes independence between the sources

Table 3 – Capture/recapture estimation of completeness on combination of suspected two source dependencies. Data obtained from patients in 2001–2005.

ICD-10	Site	True number	C: P versus D[a]	P: C versus D[b]	D: C versus P[c]	Completeness in percentage
C00–96	All sites	119,049		weighted average		98.84
C00	Lip	373	+	+		99.93
C01–02	Tongue	389	+	+		*99.85*
C03–06	Mouth, other	404	–	+	–	99.92
C07–08	Salivary glands	199	–	–	–	99.92
C09–14	Pharynx	588	–	–	+	99.12
C15	Oesophagus	934	–	–	+	97.13
C16	Stomach	2,833	–	+	+	*99.85*
C17	Small intestine	510	–	+	+	99.56
C18	Colon	10,954	–	+	+	99.82
C19–21	Rectum, rectosigmoid, anus	6,051	–	+	+	99.89
C22	Liver	627	–	+	+	98.60
C23–24	Gallbladder, bile ducts	681	–	+	+	98.39
C25	Pancreas	3,161	–	–	+	95.73
C26	Other digestive organs	438	–	+	+	96.93
C30–31	Nose, sinuses	201	+	+	+	99.85
C32	Larynx, epiglottis	609	–	–	+	98.55
C33–34	Lung, trachea	11,235	–	–	+	96.91
C38	Mediastinum, pleura	100	–	+	+	98.47
C40–41	Bone	216	–	–	+	97.12
C43	Melanoma of the skin	5,255	–	+	+	99.76
C44	Skin, non-melanoma	6,040	–	+	+	99.78
C45	Mesothelioma	360	–	+	+	99.78
C46	Kaposi's sarcoma	46	–	+	+	96.62
C47	Autonomic nervous system	47	+	+		
C48–49	Soft tissues	625	–	+	–	99.71
C50	Breast	13,701	–	+	+	99.95
C53	Cervix uteri	1,469	–	+	+	99.97
C54	Corpus uteri	3,160	–	+	+	99.98
C55	Uterus, other	46	–	+	+	97.81
C56	Ovary	2,270	–	+	+	99.95
C51–52, C57	Other female genital	662	–	+	+	99.82
C58	Placenta	20	–	+		
C61	Prostate	16,559	–	+	+	99.75
C62	Testis	1,280	–	+	+	99.92
C60, C63	Other male genital	220	+	+	+	99.66
C64, C65	Kidney including renal pelvis	3,101	–	+	+	98.96
C66–68	Bladder, ureter, urethra	6,270	–	+	+	99.67
C69	Eye	308	+	–	–	98.76
C70–72	Central nervous system	4,572	–	–	+	93.81
C73	Thyroid gland	1,000	–	+	+	99.81
C37, C74–75	Other endocrine glands	118	–	+	+	98.61
C39, C76, C80	Other or unspecified	2,614	–	+	+	97.03
C81	Hodgkin lymphoma	561	–	+	+	99.93
C82–85, C96	Non-Hodgkin lymphoma	3,663	–	–	+	98.61
C88	Malignant immunoproliferative diseases	221	–	+	+	98.42
C90	Multiple myeloma	1,680	–	+	+	95.45
C91–95	Leukaemia	2,678	–	–	+	94.64

Direction of dependency:
aC: P versus D: between pathology and death certificate sources for all clinical records.
bP: C versus D: between clinical and death certificates sources for all pathology records.
cD: C versus P: between clinical and pathology sources for all death certificates.
Shading indicates which of the two sources were merged.
The two sources with same-direction dependencies were collapsed to a single group and compared with the third (other-direction dependant) source in a two-source capture-recapture analysis to obtain the estimates of completeness. Italicised completeness estimates should be interpreted with particular caution: either they were derived for sites where the numbers of patients from the death certificate source were relatively sparse (C00, C01-02), or dependencies were in the same direction for all three two-source estimates (C07-08, C30-31 and C60,63). In the latter instance, the two sources with the strongest dependencies were collapsed, and completeness estimated using the same procedure as above. Completeness estimates for C47 and C58 could not be obtained due to low number of cases.

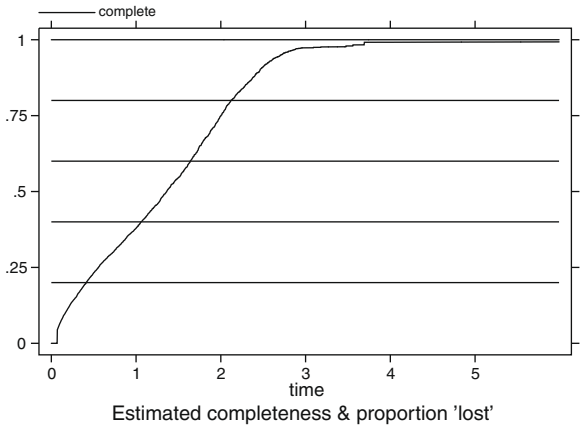

Fig. 5 – Completeness of ascertainment according to time since diagnosis, all cancer sites combined, in Norway, based on registrations in 1999.

of information, although Brenner has shown that the method often gave reliable estimates when two sources were evaluated against the third source.[25] Among other assumptions, the flow method requires the date of the first registration of each case is systematically recorded, and remain unaltered, irrespective of further data acquisition. It is difficult under such circumstances to decipher which of the methods would yield the more reliable estimates of completeness. A thorough comparison between the two methods for specific cancer types is certainly warranted, with a focus on the similarities and differences between the completeness estimates in relation to the specific assumptions underlying each method.

The semi-quantitative assessments of completeness further supported the concept of close-to-complete incidence data; the limited fluctuation in the time trends of cancer incidence rates, for instance, indicates stable reporting routines with time. The childhood cancer incidence rates were above the lower limit of the reference intervals, indicating that under-reporting is not likely. For girls aged 5–9 years, the Norwegian rates were above the upper limits of the reference interval. This could be an indication of duplicate registration, but in light of the identification of patients by their unique 11-digit personal identification number, combined with the name and address, this interpretation is unlikely to be correct.

Observed M:I ratios can be compared with standard values from the same region and tests performed to determine significant differences (see Appendix, Parkin and Bray, 2008b). The method was used in this paper to compare the M:I ratios of 34 cancers in Norway with those in Finland under the assumption that the latter country was applicable as a standard and death registration was complete. Finland is well known to have high quality data and a high degree of completeness and would seem reasonable to consider as such a reference.[26] While 11 of the 34 cancer sites analysed were flagged as statistically significant, factors other than true differences in the levels of completeness may be in operation, including variations in case fatality. Moreover, several of those cancer sites flagged were associated with few deaths (lip, penis), while others were heavily influenced by variations in diagnostic intensity and registration practices (prostate and bladder cancer, respectively).

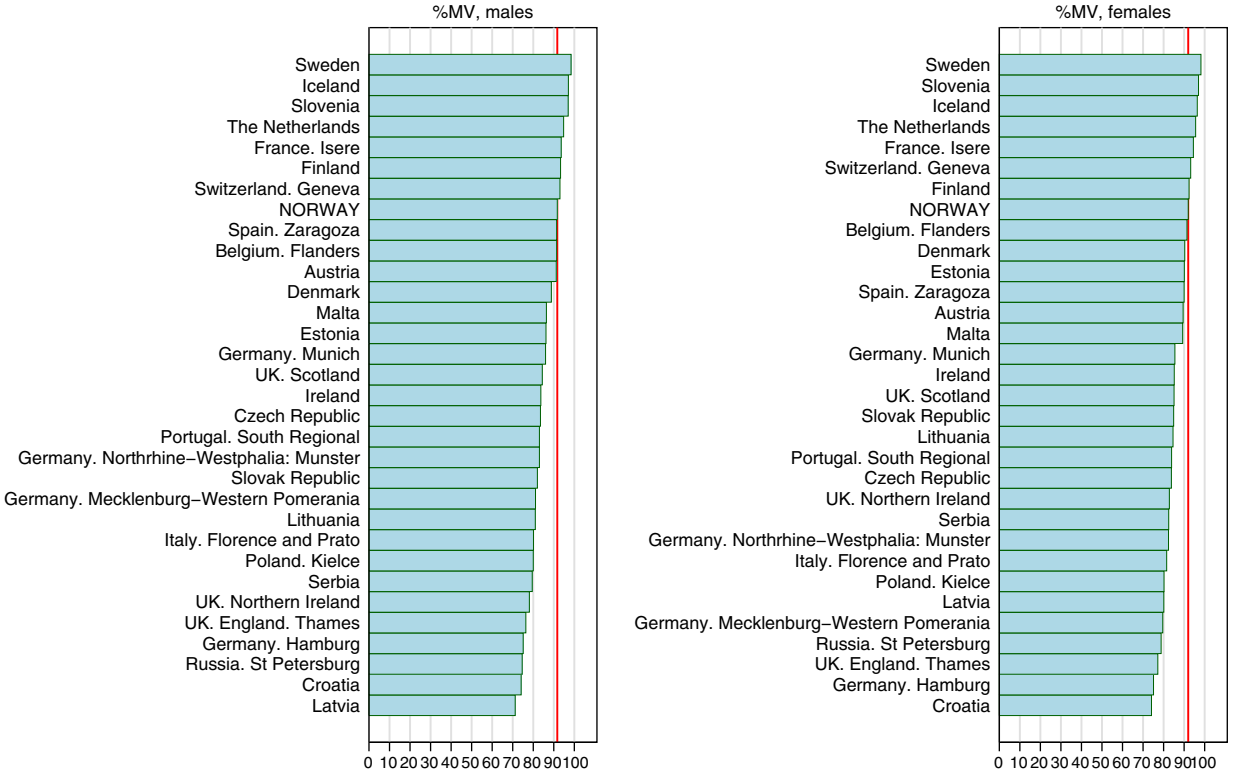

Fig. 6 – Percentage morphologically verified (%MV) cases: comparison of Norway with selected European registries for diagnoses in 1998–2002, all cancer sites combined, by sex.

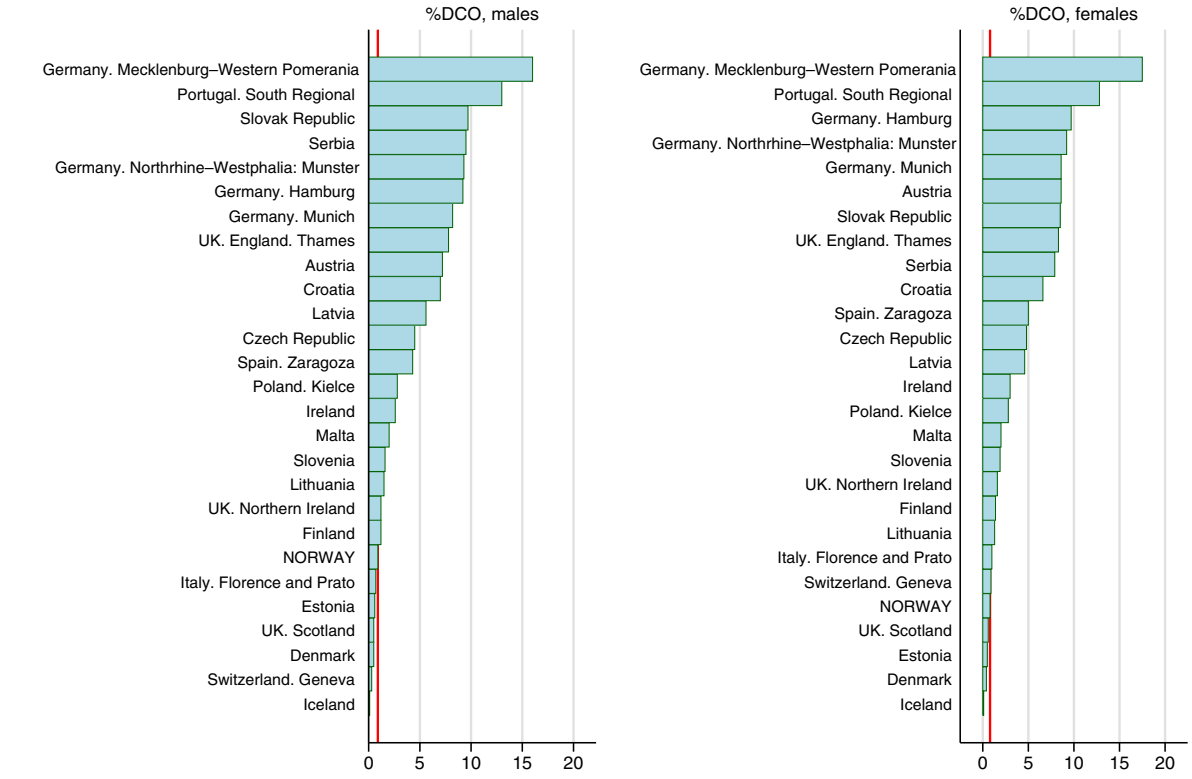

Fig. 7 – Percentage death certificate only (%DCO): comparison of Norway with selected European registries for diagnoses in 1998–2002, all cancer sites combined, by sex.

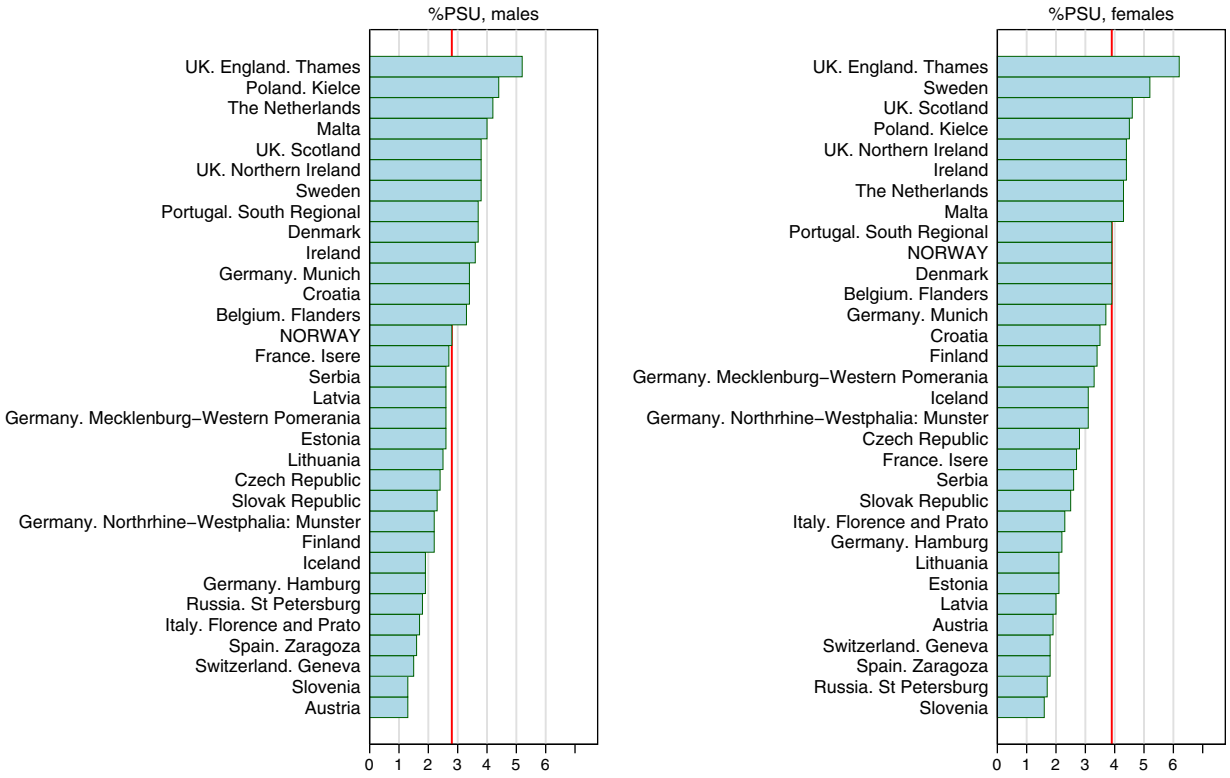

Fig. 8 – Percentage primary site unspecified (%PSU): comparison of Norway with selected European registries for diagnoses in 1998–2002, all cancer sites combined, by sex.

Table 4 – Registered cancer cases in Norway, 2005 as obtained from the registry extracted 29th November 2006 and 19th November 2007.

ICD-10	Site	Cases 2005		Difference	%
		2006	2007		
C00–96	All sites	24,199	24,730	531	2.2
C00	Lip	77	85	8	10.4
C01–02	Tongue	77	76	–1	–1.3
C03–06	Mouth, other	81	81	0	0.0
C07–08	Salivary glands	48	51	3	6.3
C09–14	Pharynx	132	127	–5	–3.8
C15	Oesophagus	198	197	–1	–0.5
C16	Stomach	515	531	16	3.1
C17	Small intestine	89	91	2	2.2
C18	Colon	2219	2249	30	1.4
C19–21	Rectum, rectosigmoid, anus	1228	1251	23	1.9
C22	Liver	131	136	5	3.8
C23–24	Gallbladder, bile ducts	140	146	6	4.3
C25	Pancreas	575	617	42	7.3
C26	Other digestive organs	107	97	–10	-9.3
C30–31	Nose, sinuses	38	40	2	5.3
C32	Larynx, epiglottis	109	113	4	3.7
C33–34	Lung, trachea	2252	2313	61	2.7
C38	Mediastinum, pleura	22	18	–4	–18.2
C40–41	Bone	43	42	–1	–2.3
C43	Melanoma of the skin	1136	1151	15	1.3
C44	Skin, non-melanoma	1342	1335	–7	–0.5
C45	Mesothelioma	73	79	6	8.2
C46	Kaposi's sarcoma	9	11	2	22.2
C47	Autonomic nervous system	9	10	1	11.1
C48–49	Soft tissues	124	130	6	4.8
C50	Breast	2798	2817	19	0.7
C53	Cervix uteri	292	297	5	1.7
C54	Corpus uteri	666	669	3	0.5
C55	Uterus, other	11	7	–4	–36.4
C56	Ovary	422	424	2	0.5
C51–52, C57	Other female genital	153	150	–3	–2.0
C58	Placenta	6	6	0	0.0
C61	Prostate	3631	3665	34	0.9
C62	Testis	249	255	6	2.4
C60, C63	Other male genital	48	50	2	4.2
C64	Kidney excluding renal pelvis	565	581	16	2.8
C65	Renal pelvis	51	51	0	0.0
C66–68	Bladder, ureter, urethra	1227	1245	18	1.5
C69	Eye	47	50	3	6.4
C70–72	Central nervous system	806	953	147	18.2
C73	Thyroid gland	226	227	1	0.4
C37, C74–75	Other endocrine glands	19	23	4	21.1
C39, C76, C80	Other or unspecified	504	496	–8	–1.6
C81	Hodgkin lymphoma	113	115	2	1.8
C82–85, C96	Non-Hodgkin lymphoma	780	775	–5	–0.6
C88	Malignant immunoproliferative diseases	32	35	3	9.4
C90	Multiple myeloma	330	360	30	9.1
C91–95	Leukaemia	449	502	53	11.8

The M:I ratio analyses did, however, flag both multiple myeloma and other leukaemia, and reiterates previous findings of under-reporting of haematological malignancies at the Registry.[2] The registration of multiple myeloma and leukaemia were 78.6% and 91.9% complete, respectively,[2] in line with the reports from many other countries.[27–29] Lund explained the under-reporting partly as a result of poor reporting routines in the departments of internal medicine. A recent study of the quality of registration of myeloma in Norway found a completeness of 92.7% in the 1990s, based on an independent case ascertainment method.[30] This corresponds well with our estimate of 95.5% completeness based on the capture–recapture method, indicating that the completeness has increased since the 1970s, but still not at a level seen for most solid tumours.

The diagnosis of haematological malignancies is often given by a clinician, partly based on the haematology report. The possibility to send out reminders to the clinicians, based on Norwegian hospital discharge files since the late 1990s, has probably contributed significantly to the increase

in completeness. This trace-back system collects cases with a cancer code (ICD-10) from hospital records, and has probably also increased the level of completeness for cancers with a low proportion of morphologically verified tumours, such as cancers of the central nervous system and pancreas. A large proportion of brain tumours are diagnosed based on imaging, without morphologically verification. Before the hospital discharge files were used to send out reminders in Norway, some of these tumours were probably not registered at all, or were only picked up as death certificate initiated cases.

During the recent years, the Cancer Registry has extended beyond collecting information at the time of diagnosis, to include more clinical data on treatment and follow-up of patients.[31–33] This has been accompanied with feedback to each hospital on various quality indicators, including the proportion of patients with local recurrence and receiving a magnetic resonance imaging, to enable benchmarking of the hospital with the rest of the country.[34] This system calls for the timely publication of cancer statistics. The Norwegian cancer incidence for 2005 was published by the end of 2006, with a calculated under-reporting of 2.2% of cases that arrived after the date of first publication. The late arrival of clinical notifications for some cancers, like those of the lip, central nervous system and leukaemia, causes deficits of a larger magnitude (10–20%). For the CRN, rapid dissemination of accurate and close-to-complete data outweighs concerns regarding the accrual of additional cases over several years, given that they represent a minor proportion of the total annual incidence. Timeliness is expected to further improve in the coming years, due to the recent introduction of electronic data reporting from the hospitals.

Validity in the present study was evaluated according to three indicators. For all sites combined, 93.8% of the cases were morphologically verified (MV), 0.9% were registered on the basis of a death certificate only (DCO), and 2.2% were registered with primary site unknown (PSU). Sub-analyses showed that the proportion of PSU was twice as high in cancer patients aged 75 years or more, compared to patient aged between 65 and 74 years. The pattern of decreased validity of data for the elderly patients are also reflected in a study of the elderly population in the Netherlands, where unknown primary tumour site was ranked as one of the most frequent site-specifications in older age-groups.[35]

Comparisons with selected European registries 1998–2002 from CI5 Volume IX[10] showed that Norway had the eighth highest %MV of thirty two countries, suggesting a relative high validity of the registered cases. The accuracy of registry data in Norway has been reviewed for cancers of the central nervous system (CNS),[36] cancers of the head and neck,[4] prostate[3] and ovary.[5] All these studies assessed accuracy by re-examination of a random data sample. The evaluation of the CNS found that the rate of errors concerning the overall incidence rates was 0.3%, whereas in total, 9.3% of the cases evaluated had errors of varying severity.[36] The re-evaluation of 15992 head and neck cancers cases, reported that 1.6% of these cases were registered with errors that needed to be corrected. The level of erroneous coding was, however, highest in the late 1960s and early 1970s.[4] The evaluation of accuracy in prostate cancer registration, found that 0.5% of the data elements had errors, and a detailed review of 1% of the registered cases diagnosed in 1957–1989, revealed that one of 298 cases (0.3%) was erroneously registered as a cancer case.[3] Finally, a study assessing the validity of ovarian cancer, found the accuracy of diagnosis to be around 92%.[5]

These quality investigations above tended to be part of larger studies. While relatively few studies have been published on data quality specifically in Norway, researchers perform quality checks including searches for errors and inconsistencies, extreme values and rechecks of unexpected outcomes, before analyses are conducted. Detected errors are reported to the coding staff for correction in the database. Thus, as the research activity has increased within the Registry during the recent years, there is a more thorough evaluation of data quality extending beyond the routine checks employed by the coding staff, improving the accuracy of the data at the CRN. Further improvements in data quality may be achieved by the establishment of the cancer-specific clinical registries.[31–33] At present, registries for prostate, colorectal, lung and ovarian cancer and malignant melanoma have been launched, with lymphoma and breast cancer registries planned to begin in 2009.

6. Conclusion

This review indicated that the routines in place at the CRN yield comparable data that can be considered reasonably accurate, close-to-complete and timely, and serves as a justification for our policy of reporting annual incidence one year after the close of registration.

Conflict of interest statement

None declared.

Authors contribution

Study concept and design: FB, DMP, BM
Acquisition of data: IKL
Analysis and interpretation of data: IKL, MS
Drafting of the manuscript: IKL, BM, MS, TBJ
Critical revision of the manuscript for important intellectual content: IKL, MS, TBJ, FL, DMP, FB, BM.

Acknowledgements

We would like to thank the following staffs at the Cancer Registry of Norway for discussion and comment: Aage Johansen, Bjørge Sæther, Svein Erling Tysvær, Tove Dahl, Siri Larønningen, Olaug Talleraas and Christine Mellem. This study was undertaken as part of the Cancer Control using Population-based Registries and Biobanks (CCPRB) project, funded by the EU Sixth Framework Programme (FP6-2002-LifeSciHealth Contract No. 503465).

Appendix

Site	Quality aspects	Study period	Methods	Main results and conclusions	Reference
All cancer sites	Completeness	1976	Completeness was estimated by matching of cases registered in the Cancer Registry of Norway against those registered within the Economic and Medical Information system. The study included cancer cases diagnosed in 1976 in the hospitals of either Rogaland or Oppland county	The overall completeness, three years after diagnosis, was estimated to be 97.6%. In total, 34 cases of 1417 cancer cases registered in EMI were not registered in the Cancer Registry of Norway. The study revealed local differences in reporting routines between the hospital departments. For cancer of lymphatic and haemopoietic tissues reporting was markedly poorer than that for other sites; especially for multiple myeloma or leukaemia, where completeness was estimated to be 78.6% and 91.8%, respectively	[2]
Neoplasms of central nerve system	Validity/quality of data	1955–1984	All cases registered with primary central nervous system neoplasms diagnosed in the period between 1955 and 1984 were included (n = 8933)	Validity. In total, 86 cases were found to be registered with minor or major errors that were corrected. The main source of error was misinterpretation of data by cancer registry staff (67 of 109 cases)	[36]
			Validity. The dataset was checked for errors, and seven categories of neoplasms records likely to be faulty were defined and identified for extensive study (n = 109)	Quality of data. Errors concerning total incidence rates and rates for main groups of gliomas, meningiomas and neurilemmomas represented 0.3% of the total, and altogether 9.3% of errors of varying severity were revealed.	
			Quality of data. The second approach was to evaluate the quality of the corrected data. This was done by a random draw of 300 cases	The study concluded that the data in the Norwegian Cancer Registry was of sufficient validity	
Prostate cancer	Completeness	Completeness 1960, 1975 and 1981	Completeness. Data were matched and evaluated against diagnostic indices at eight selected hospitals in Norway and against death certificates from Statistics Norway	Completeness. The deficiency in reporting of prostate cancer was less than 1%. The completeness was found to improve with the number of years following registration: 99% and close to 100% after 5 and 10 years, respectively. The completeness following one year of registration was 95%	[3]
	Internal validity	Internal validity 1957–1986	Internal validity. The validity control was based on a detailed reanalysis of an approximately 1% sample of the registered data during the period 1957–1986	Internal validity. The validity control revealed errors in 0.5% of the data elements. One false positive case was found among 298 cases controlled (0.3%)	

(*continued on next page*)

Appendix (*continued*)

Site	Quality aspects	Study period	Methods	Main results and conclusions	Reference
Head and neck cancer	Consistency of classification and coding and basis of diagnosis	1953–1991	Consistency and basis of diagnosis. All histological codes for (16,104) cases of head and neck malignancies were reviewed. A set of criteria for acceptable histological diagnosis were defined, and cases that did not meet the acceptance criteria were subjected to a re-evaluation of the pathologists' reports (369 cases). The percentage of cases not histologically verified or solely registered on the basis of death certificate was calculated for consecutive decades throughout the study period	Consistency and basis of diagnosis. Of 16,104 cases, 369 (2.3%) fulfilled criteria for further scrutiny. No changes were made in 98.4% of the cases in the crude series. The level of erroneous coding was highest in the late sixties and the early seventies, and dropped to an all time low after 1980. Only 46 cases were registered solely on the basis of death certificates. The percentage of cases lacking microscopic confirmation dropped from 5.7% in 1953–1962 to 2.1% in 1983–1991	[4]
	Completeness		Completeness. Completeness was checked against hospital-based registries in two university clinics in Oslo	Completeness. All patients in the hospital-based patient registry were also registered in the Cancer Registry. In addition, 44 patients treated in the hospitals were found in the cancer registry and not in the hospital registry	
Ovarian	Completeness	1987–1996	Completeness. Re-ascertainment. Persons identified with ovarian cancer in the Cancer Registry or in the hospital's discharge registers were included in the study	Completeness. The overall completeness for ovarian cancer was found to be 99.6%. The organ specific completeness of registration of histologic verified ovarian cancer within the Cancer Registry was 95.3%; 0.9% was erroneously coded and 3.5% had their diagnosis changed to ovarian cancer at re-evaluation	[5]
	Reproducibility of the histopathologic diagnosis		Histological re-evaluation of 729 cases	Histological re-evaluation. In total 91% of the cases had a primary histological diagnosis. The accuracy of the diagnosis for 591 cases identified with a histological diagnosis was estimated at 92%. Coding errors were found in 2% of these cases, while in 6% of the cases it was not possible to reproduce the original diagnosis of ovarian cancer at re-evaluation	

REFERENCES

1. Det Kongelige Helsedepartement. Forskrift om innsamling og behandling av helseopplysninger i Kreftregisteret (Kreftregisterforskriften) av 21. desember 2001 nr 1477.
2. Lund E. Pilot study for the evaluation of completeness of reporting to the cancer registry. In: *Incidence of cancer in Norway 1978*. Oslo: Cancer Registry of Norway; 1981. p. 11–4.
3. Harvei S, Tretli S, Langmark F. Quality of prostate cancer data in the cancer registry of Norway. *Eur J Cancer* 1996;**32A**:104–10.
4. Mork J, Thoresen S, Faye-Lund H, Langmark F, Glattre E. Head and neck cancer in Norway. A study of the quality of the Cancer Registry of Norway's data on head and neck cancer for the period 1953–1991. *APMIS* 1995;**103**:375–82.
5. Tingulstad S, Halvorsen T, Norstein J, Hagen B, Skjeldestad FE. Completeness and accuracy of registration of ovarian cancer in the cancer registry of Norway. *Int J Cancer* 2002;**98**:907–11.
6. Larsen IK, Småstuen M, Parkin DM, Bray F. Data Quality at the Cancer Registry of Norway. In: *Cancer in Norway 2006 – Cancer incidence, mortality, survival and prevalence in Norway*. Oslo: Cancer Registry of Norway; 2007.
7. Bray F, Parkin DM. Evaluation of data quality in the cancer registry: Principles and methods. Part I. Comparability, Validity and Timeliness. *Eur J Cancer* 2009;**45**:747–55.
8. Parkin DM, Bray F. Evaluation of data quality in the cancer registry: Principles and methods. Part II. Completeness. *Eur J Cancer* 2009;**45**:756–64.
9. Parkin DM, Whelan SL, Ferlay J, Teppo L, Thomas DB. *Cancer incidence in five continents, Vol. VIII (IARC Scientific Publications No. 155)*. Lyon: IARC; 2002.
10. Curado MP, Edwards B, Shin HR, et al. *Cancer Incidence in Five Continents, vol. IX (IARC Scientific Publication No. 160)*. Lyon: IARC; 2007.
11. *Manual of the international statistical classification of diseases, injuries, and causes of death*. Geneva: World Health Organization; 1948.
12. Task force to revise the Manual of tumor nomenclature and coding for the American Cancer Society IE. *Manual of tumor nomenclature and coding: 1968 ed*. New York: American Cancer Society, Inc.; 1968.
13. *Manual of the international statistical classification of diseases, injuries and causes of death: based on the Recommendations of the Seventh Revision Conference, 1955, and adopted by the Ninth World Health Assembly under the WHO Nomenclature Regulations*. Geneva: World Health Organization; 1957.
14. *International classification of diseases for oncology: ICD-O*, 2nd ed. Geneva: World Health Organization; 1990.
15. Lennert K. *Histopathology of non-Hodgkin's lymphomas: (based on the Kiel Classification)*. Berlin: Springer; 1981.
16. Fritz A, Percy C, Andrew J, et al., editors. *International classification of diseases for oncology: ICD-O*. 3rd ed. Geneva: World Health Organization; 2000.
17. *International statistical classification of diseases and related health problems*. Geneva: World Health Organization; 1992.
18. *Standards and guidelines for cancer registration in Europe: The ENCR recommendations*, vol. 1. Lyon: International Agency for Research on Cancer; 2003.
19. Kyllonen LE, Teppo L, Lehtonen M. Completeness and accuracy of registration of colorectal cancer in Finland. *Ann Chir Gynaecol* 1987;**76**:185–90.
20. Barlow L, Westergren K, Holmberg L, Talback M. The completeness of the Swedish Cancer Register – a sample survey for year 1998. *Acta Oncol* 2008;1–7.
21. Thorsteinsson R, Sorensen M, Jensen TL, et al. Central nervous system tumours in children. An evaluation of the completeness and validity of the Cancer Registry. *Ugeskr Laeger* 2005;**167**:3782–5.
22. 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.
23. Mattsson B, Wallgren A. Completeness of the Swedish Cancer Register. Non-notified cancer cases recorded on death certificates in 1978. *Acta Radiol Oncol* 1984;**23**:305–13.
24. Storm HH. Completeness of cancer registration in Denmark 1943–1966 and efficacy of record linkage procedures. *Int J Epidemiol* 1988;**17**:44–9.
25. Brenner H. Use and limitations of the capture-recapture method in disease monitoring with two dependent sources. *Epidemiology* 1995;**6**:42–8.
26. Teppo L, Pukkala E, Lehtonen M. Data quality and quality control of a population-based cancer registry. Experience in Finland. *Acta Oncol* 1994;**33**:365–9.
27. Phekoo K, Moller H, Richards M, Schey S. Comparison of a specialist haematological malignancy database against a regional cancer registry: Case ascertainment and diagnostic accuracy. *Brit J Haematol* 2002;**119**:697–705.
28. Astrom M, Bodin L, Tidefelt U. Adjustment of incidence rates after an estimate of completeness and accuracy in registration of acute leukemias in a Swedish population. *Leuk Lymphoma* 2001;**41**:559–70.
29. Norgaard M, Skriver MV, Gregersen H, et al. The data quality of haematological malignancy ICD-10 diagnoses in a population-based hospital discharge registry. *Eur J Cancer Prev* 2005;**14**:201–6.
30. Aasbrenn M, Langmark F, Wisløff F. The quality of multiple myeloma registration in the Norwegian Cancer Registry [in Norwegian]. *Tidsskr Nor Laegeforen*. in press.
31. Strand TE, Rostad H, Moller B, Norstein J. Survival after resection for primary lung cancer: A population based study of 3211 resected patients. *Thorax* 2006;**61**:710–5.
32. Wibe A, Moller B, Norstein J, et al. A national strategic change in treatment policy for rectal cancer – implementation of total mesorectal excision as routine treatment in Norway. A national audit. *Dis Colon Rectum* 2002;**45**:857–66.
33. Paulsen T, Kaern J, Kjaerheim K, Haldorsen T, Trope C. Influence of interval between primary surgery and chemotherapy on short-term survival of patients with advanced ovarian, tubal or peritoneal cancer. *Gynecol Oncol* 2006;**102**:447–52.
34. Dørum L, Gaard M, Vonen B. Sykehusrapporten. Behandling av pasienter med endetarmskreft i Norge. Resultater fra perioder 1993–2004 [in Norwegian]. Kreftregisteret; 2008.
35. de Rijke JM, Schouten LJ, Hillen HF, et al. Cancer in the very elderly Dutch population. *Cancer* 2000;**89**:1121–33.
36. Helseth A, Langmark F, Mork SJ. Neoplasms of the central nervous system in Norway. I. Quality control of the registration in the Norwegian Cancer Registry. *APMIS* 1988;**96**:1002–8.