Original Articles

Accuracy of ovarian cancer ICD-10 diagnosis in a Danish population-based hospital discharge registry

M.S. Tetsche[1,2], *M.D.*; **M. Nørgaard**[1,3], *M.D.*; **M.V. Skriver**[1], *Cand. Scient.*; **E.S. Andersen**[2], *M.D., DMSc.*; **T.L. Lash**[4], *DSc., MPH.*; **H.T. Sørensen**[1], *Prof., M.D., DMSc.*

[1]Department of Clinical Epidemiology, Aarhus University Hospital, [2]Department of Gynaecology, Aalborg Hospital, Aarhus University Hospital (Denmark); [3]Department of Medicine, Vanderbilt University Medical Center, Vanderbilt-Ingram Cancer Center, Nashville, TN; [4]Department of Epidemiology, Boston University School of Public Health, Boston, MA (USA)

Summary

Objective: We estimated the accuracy of ICD-10 diagnosis of ovarian cancer in a Danish discharge registry (HDR) by comparing it with Cancer Registry data (DCR).

Study design and setting: Patients (N = 489) living in North Jutland County, Denmark with ovarian cancer or borderline tumour registered in the HDR or the DCR. We estimated the completeness and positive predictive value (PPV) of ovarian cancer discharge diagnosis. Mortality rates were constructed for both registries.

Results: The completeness in the HDR for ovarian cancer was 96% (95% confidence interval [CI]: 94%-98%) and PPV was 87% (95% CI: 85%-90%). 87 (18%) of the patients coded with ovarian cancer in the HDR had borderline tumours. When borderline tumours were excluded from the DCR, the PPV declined to 69% and the completeness did not change.

The mortality rate ratio for ovarian cancer registered in the HDR compared to the DCR was 1.08 (95% CI: 0.90-1.29).

Conclusion: The discharge data (ICD-10) had some misclassification, but can be a valuable tool in assessment of the prognosis of ovarian cancer.

Key words: Epidemiology; Ovarian neoplasm; ICD-10 diagnosis; Data quality.

Introduction

Ovarian cancer is the seventh most frequent cancer in women worldwide and the Scandinavian countries exhibit some of the highest incidence rates [1]. Denmark has about 600 cases of ovarian cancer diagnosed annually [2]. Ovarian cancer is the leading cause of death from gynaecologic malignancies in most of the Western world. The survival, however, varies between the different countries, and Denmark has a poorer relative five-year survival (32% of incident cases) compared to other developed countries [3].

To support evidence-based practice and to track the success of ovarian cancer care, long-term population-based outcome data are needed. Administratively, clinical databases are increasingly being used for the purposes of examining and monitoring health outcomes. Continuously available data on survival is an important tool for improving survival after ovarian cancer care by modifying therapeutic strategies in reaction to those with the best demonstrated effectiveness. The benefits of using such databases are that they already exist, are inexpensive, and in particular are updated daily [4]. The users of registry data have, however, no control of how the data are collected and the quality of the data [4, 5], and it is well known that the validity of administrative data varies across diseases [6].

The Danish Hospital Discharge Registries (HDR) are available in each hospital. An earlier Danish study of the quality of ovarian cancer ICD-8 discharge diagnosis showed a questionable quality of the data in the HDR [7]. However, it has been expected that implementation of ICD-10 in the HDR would improve the data validity. Thus, it would be of great value to characterize the validity of the HDR data, so medical departments could continuously monitor the effectiveness of therapeutic outcomes [8]. Further, it would be an important tool for physicians in their daily clinical work to have quality assurance systems based on recently updated data sources [9, 10].

We therefore estimated the accuracy of the ICD-10 ovarian cancer discharge diagnosis in the population based North Jutland Hospital Discharge Registry by comparing it with data from the Danish Cancer Registry and the County Pathology Registry in the period of 1994-1999. We also compared the ovarian cancer survival among the cases recorded in the North Jutland County Hospital Discharge Registry with the survival among the cases recorded in the Danish Cancer Registry for residents of North Jutland County.

The study obtained support from the Western Danish Research Forum for Health Sciences and Leo and Ingeborg Dannin Foundation for Scientific Research.

Revised manuscript accepted for publication November 17, 2004

Materials and Methods

We conducted the study in Denmark's North Jutland County – which has approximately 500,000 inhabitants – from January 1, 1994 to December 31, 1999.

The North Jutland Hospital Discharge Registry

Every Danish county has a hospital discharge registry from which data are transferred to the nationwide Hospital Discharge Registry (HDR). The Danish HDR has been operating since 1977 and covers all patients admitted to hospitals in Denmark for somatic diseases. For each hospitalization, the files of the HDR include date of admission and discharge, any surgical procedure performed, and up to 20 discharge diagnoses.

Between 1977 and 1993 a Danish version of the International Classification of Diseases (ICD) 8th revision had been used [11]. Since 1994 data have been classified according to the ICD 10th revision.

To identify the patients for our study, we used the ICD-10 codes C56.0-C56.9 for ovarian cancer. Ovarian tumours include both borderline tumours and invasive tumours. However, there is no specific code for borderline tumours in the ICD-10 classification, so it is not possible to separate these from invasive tumours. The procedure in many departments is to classify ovarian borderline tumours as ovarian cancer in ICD-10, thus borderline tumours are included in the study. To exclude the prevalent cases of ovarian cancer diagnosed in the period 1977-1993 we used ICD-8 codes 183.00-183.09. ICD-9 has never been used in Denmark for the HDR.

The Danish Cancer Registry

The Danish Cancer Registry (DCR) is a population-based registry containing data on the incident cases of cancer throughout Denmark since 1943. In 1987 notification by physicians of malignant and related diseases to the DCR was made compulsory in Denmark. Since then the cancer cases have also been identified through linkage to the HDR and the Danish Registry of Causes of Death, so that missing reports can be included. This process has to ensure the completeness of the DCR, but it causes a certain delay of data availability. Furthermore, the data are not available for the specific medical departments. All DCR data are reclassified to the modified ICD-7. We used ICD-7 codes 175.0, 175.1, 175.2, 175.3, 375.0, 475.0, 775.5, 875.0 and. 875.5 for ovarian cancer. Ovarian borderline tumours are registered separately in the DCR, so to identify ovarian borderline tumours we used ICD-7 codes 575.0, 575.1, 575.2 and 575.5.

The County Pathology Registry

The Department of Pathology, Aalborg Hospital, has computerized records of histopathology and cytology for all pathologically confirmed cases of cancer in the county. The record includes the date of diagnosis. We reviewed the pathological records on the patients identified in only one of the two registries, if pathological records were available.

Histopathological findings were available for 85% of the patients; on-line discharge summaries were available for a further 14% of the patients. In 1% of the patients, we found no information. The registry diagnosis was classified as correct if the pathological diagnosis was ovarian cancer, as most likely correct if the diagnosis was based on clinical suspicion in a patient with large tumour masses in the pelvis or based on inconclusive histopathology with carcinoma cells, and as incorrect if the pathological diagnosis was clearly not ovarian cancer.

Record linkage between data sources

Use of the 10-digit personal identification number, which was assigned to all Danish citizens in 1968 and has been assigned to each citizen at birth or immigration since 1968, allowed an unambiguous record linkage between the registries. The Civil Registration System contains information on vital status, date of death for decedents, and residence of the citizens. In this study we only included the patients residing in the county of North Jutland at the time of the ovarian cancer diagnosis.

Statistical analysis

We used the data in the DCR as the standard reference. The County Pathology Registry could not be used as the standard reference because all incident cases of ovarian cancer could not be identified in this registry. Two measures are used to validate registries: completeness and positive predictive value (PPV) [4]. To estimate the completeness of registration of patients, we calculated the proportion of patients registered with ovarian cancer in the DCR who were also found with ovarian cancer diagnoses in the HDR [4]. The numerator was the number of patients registered in both registries and the denominator was the number of all the patients registered in the DCR.

We defined PPV as the proportion of patients registered with ovarian cancer in the HDR who also had ovarian cancer diagnoses in the DCR [4]. The numerator was the number of patients registered in both registries and the denominator was the number of all patients registered in the HDR.

We chose to perform analyses with and without borderline tumours included in data from the DCR. Calculations of completeness and PPV were also stratified by age group (< 50 years, 50-75 years and > 75 years). All estimates are presented with 95% confidence interval (95%CI) [12].

The misclassification in the HDR and the DCR was calculated by counting the patients who did not have an ovarian cancer diagnosis in the pathological reviews and compared it to all the patients in the HDR and the DCR, respectively.

Survival

To compare the survival of ovarian cancer in the two registries we constructed Kaplan-Meier curves for both registries. Cox proportional hazards regression analyses were used to compare survival estimates for patients in the HDR and in the DCR, providing an estimate of the mortality rate ratio (MR). Curves of one year-survival rates over the years of follow-up were constructed for both registries to compare the one-year survival. Estimates are presented with 95% confidence interval (95% CI). The analyses were performed in SAS® System, 8.2.

Results

In the period 1994-1999 we identified 489 patients in the HDR and the DCR, 411 (84%) were found in both registries, 59 (12%) were found only in the HDR and 19 (4%) were found only in the DCR. We found 87 (18%) of the patients coded with ovarian cancer in the HDR actually had borderline tumours.

Degree of completeness and PPV

The completeness was 96% (95%CI: 94%-98%) and PPV was 87% (95%CI: 85%-90%) when we used the DCR (including borderline tumours) as the reference standard (Table 1).

Table 1. — *Number of patients with an initial diagnosis of ovarian cancer in the North Jutland Hospital Discharge Registry (HDR), Denmark, in the Danish Cancer Registry (DCR), and in both registries. Data from the DCR are calculated with and without borderline tumours. Three age groups were formed for the data including borderline tumours. Degree of completeness and positive predictive value (PPV) are given as percent.*

	Patients registered in:					
	Both registries n (%)	Only HDR n (%)	Only DCR n (%)	Total n	Degree of completeness % (95% CI)	Positive predictive % (95% CI)
DCR without borderline tumour	324 (67)	146 (30)	15 (3.1)	485	96 (93-97)	69 (65-73.0)
DCR with borderline tumour	411 (84)	59 (12)	19 (3.9)	489	96 (94-98)	87 (85-90)
DCR with borderline tumour Age < 50 year	97 (87)	7 (6.3)	7 (6.3)	111	93 (89-98)	93 (89-98)
DCR with borderline tumour 50 year < Age < = 75 year	247 (86)	35 (12)	6 (2.0)	288	99 (96-99.5)	88 (84-91)
DCR with borderline tumour Age > 75 year	65 (71)	19 (21)	8 (8.7)	92	89 (82-96)	77 (68-86)

The completeness was 96% (95%CI: 93%-97%) and PPV was 69% (95%CI: 65%-73%) when the DCR (without borderline tumours) was used as the reference standard (Table 1). Completeness did not show any substantial dependence on age group, but the PPV declined with age (Table 1).

Pathological record reviews

Data from the 78 pathological reviews (patients only registered in one of the registries) are shown in Table 2.

Table 2. — *Results of the validation of diagnoses by review of pathology records for all 78 patients (16.0%) registered only in one registry, either in the Danish Cancer Registry (DCR), n = 19, or the North Jutland Hospital Discharge Registry (HDR), n = 59.*

	Patients registered only in: DCR n (%)	HDR n (%)
Diagnosis correct	6 (32)	19 (32)
Diagnosis correct but made post-mortem	1 (5)	–
Diagnosis most likely/possibly correct*	7 (37)	13 (22)
Benign ovarian disease	1 (5)	4 (7)
Other gynaecological cancer	1 (5)	4 (7)
Non-gynaecological cancer**	1 (5)	14 (24)
No cancer at all	1 (5)	–
Diagnosis prevalent***	–	4 (7)
Unknown	1 (5)	–
Diagnosed ovarian cancer in another county	–	1 (2)
Total	19	59

*Diagnosis strongly suspected but based on clinical suspicion in a patient with large tumour masses in the pelvis or based on inconclusive histopathology with carcinoma cells.
**Colon cancer, urological cancer or breast cancer.
***Diagnosis correct but made before 1994; the diagnosis was made in another county where the women lived before 1994.

For the 19 patients registered only in the DCR, the diagnosis was judged as correct or most likely correct in 14 (74%) of the cases. For the 59 patients registered only in the HDR the diagnosis was judged as correct or most likely correct in 32 (54%) of the cases. Of the 470 patients registered with an ovarian cancer diagnosis in the HDR, 26 (5.5% [95% CI: 3.5%-7.6%]) patients could not be confirmed as having this diagnosis either by registry in the DCR or by pathological review. In the DCR the false-positive misclassification rate was 3.0% (13/430).

Of the 59 patients only registered in the HDR 32 (54%) had ovarian cancer diagnoses, 18 patients had other cancers (e.g., colon cancer, endometrial cancer or breast cancer), and all the cancers were at an advanced stage of the disease.

Survival

Survival curves based on patients registered in the HDR and in the DCR are shown in Figure 1. The Kaplan-Meier curves show slightly higher mortality rates for patients registered in the HDR, which is confirmed by the proportional hazards regression comparing the HDR patients' survival with the DCR patients' survival [MR= 1.08 (95%CI: 0.90-1.29)].

One-year survival rates over the course of the calendar period of follow-up are shown in Figure 2. For all the years shown, the one-year survival based on data from the HDR was close to the estimate based on data from the DCR.

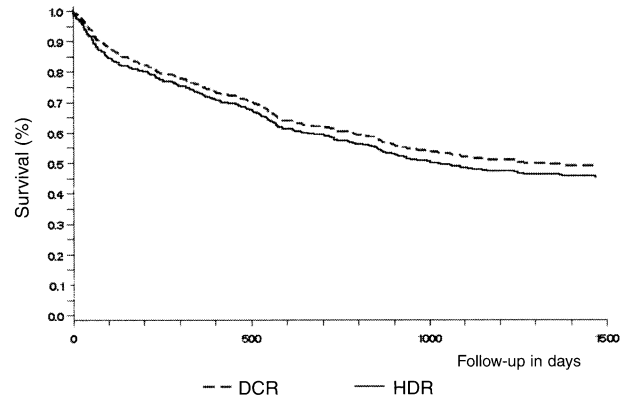

Figure 1. — Survival analysis based on data from the Hospital Discharge Registry (HDR) in North Jutland County, Denmark, compared with survival based on data from the Danish Cancer Registry in the period 1994-1999.

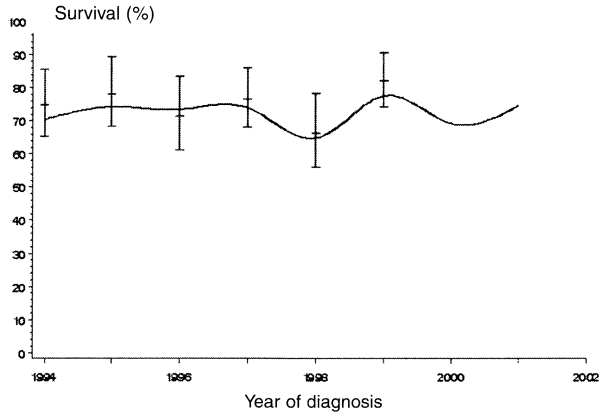

Figure 2. — One-year survival of patients with ovarian cancer. Bars represent data from the Danish Cancer Registry (DCR) with 95% CI, and the solid line represents data from the Hospital Discharge Registry (HDR) in North Jutland County, Denmark.

Discussion

We evaluated the accuracy of ovarian cancer registration in a population based hospital discharge registry and found completeness around 95% and PPV around 87% in ICD-10-coding when borderline tumours were included in the reference DCR registry. PPV declined to 69% when borderline tumours were excluded, while completeness remained approximately the same. Borderline tumours are thus a potential source of bias in the HDR. We found that the PPV was slightly higher in younger age groups than in older age groups. Although we found a minor difference in survival between patients registered in the HDR and patients registered in the DCR, the hospital discharge registry data are a possible source for outcome research and quality assessment, where the actuality of the data is important.

The optimal way to estimate the validity of administrative hospital discharge data is to measure sensitivity and specificity [4], the completeness is an estimate of the sensitivity. Since we do not know the true incidence of ovarian cancer in the entire general population, we could not estimate the specificity, but only the predictive value in the comparisons between the two registries [4].

The demand for completeness and PPV depends on the question and the design[4]. If data are used to monitor the outcomes, sensitivity and specificity should be stable over time. Since we analysed data from only a six-year time-period after the introduction of ICD-10, we did not find any clear time trend. Analyses over a longer duration may identify trends in data quality over time.

The accuracy of the coding had a small impact on survival estimates. However, the survival curves had only a minor systematic error, when we estimate survival in the HDR we overestimate the mortality by 8%. The one-year survival curve indicates that there was no major difference in survival when we compare the two registries.

There have only been a few studies in which gynaecological discharge diagnoses have been validated. A Danish study evaluated the validity of ICD-8 non-malignant gynaecological conditions and found a PPV of 78% concerning benign ovarian tumours [13]. The data quality of the HDR found in our study was slightly better than the data quality in another Danish study where ICD-8 registration of surgically treated malignant gynaecological diseases were compared to the registration in the Danish Cancer Registry [7]; they found a completeness of 83% and PPV of 90% concerning ovarian cancer (unspecified whether borderline ovarian tumours were included). Our data suggest that use of a simpler more coherent classification of diseases (ICD-10) improves the validity in comparison to the highly detailed coding (ICD-8).

Few other data exist on data quality in similar registries. A British study on ovarian neoplasms including 868 women found that only 74% of ovarian neoplasm cases were registered in the hospital admissions records (1979-1983) [14] using ICD-9 coding. A Norwegian study of 945 patients identified as having ovarian cancer in either the hospital discharge registry or the Norwegian cancer registry demonstrated a nearly 100% completeness of the Norwegian Cancer Registry when borderline tumours were excluded [15]. Not all cancer registries have high quality data. A British study including 49 ovarian cancer patients identified in a screening program of 22,000 women showed that 22% of the ovarian cancer cases were missing from the cancer registry [16].

We chose to use the DCR as the reference standard. The DCR, however, did also contain misclassified cases. We estimated an under notification of cases of 3.0%; which would lead to an underestimate of the completeness and of the PPV from the HDR. Previous Danish studies have found a deficit of 2.2% (cervical cancer) [17] and 0% (breast cancer) [18].

Hospital discharge registry data can also be used for purposes other than outcome research and quality assessment such as for identification of cases in case-control studies of risk factors for ovarian cancer. In case-control studies of serious conditions like ovarian cancer, it is important that interviews are performed soon after the diagnosis is made because of the high mortality hazard. Because the HDR is regularly updated, it is highly suitable for case-control studies. The misclassification of 6.8% found in the HDR is expected to be unrelated to information about earlier exposures, so this misclassification would ordinarily bias relative measures such as an odds ratio towards unity [4]. The bias in case-control studies can be reduced if data from the HDR are combined with the data from the County Pathology Registry or by reviewing medical records. The cases missed by the HDR would only cause bias if their exposure differed from the exposure of the registered cases [19].

If hospital discharge data are used to measure ovarian cancer incidence in a cohort design, the misclassification would probably be unrelated to how the patients were classified according to exposure. The bias of relative measures would therefore be towards unity [4].

Conclusion

Our study showed that discharge data (ICD-10) had some misclassifications; however, this had little impact on survival estimates. The accuracy of ovarian cancer discharge diagnoses in ICD-10 was slightly improved compared to ICD-8. Thus, in the future, hospital discharge data can be a valuable tool for gynaecological cancer research and for quality assessment activities. However, the researcher should bear in mind that data from the HDR can include borderline ovarian cancer cases because borderline tumours cannot be excluded.

Acknowledgements

The authors are grateful to Karsten Nielsen, M.D., Department of Pathology, Aalborg Hospital, Denmark for providing assistance with the County Pathology Registry.

References

[1] Bray F., Sankila R., Ferlay J., Parkin D.M.: "Estimates of cancer incidence and mortality in Europe in 1995". *Eur. J. Cancer,* 2002, *38*, 99.

[2] National Board of Health: "Cancer registret 2000, foreløbig opgørelse". In: Nye tal fra Sundhedsstyrelsen, 2004, 8.

[3] "Kræft i æggestokkene i tal". 1999. URL: http://www.cancer.dk/alt+om+kraeft/kraeftsygdomme/aeggestok/kraeft+i+aeggestokkene+i+tal.asp.

[4] Sørensen H.T., Sabroe S., Olsen J.: "A framework for evaluation of secondary data sources for epidemiological research". *Int. J. Epidemiol.,* 1996, *25,* 435.

[5] Baron J.A., Weiderpass E.: "An Introduction to epidemiological research with medical databases". *Ann. Epidemiol.*, 2000, *10,* 200.

[6] Sørensen H.T.: "Regional administrative health registries as a resource in clinical epidemiology. A study of options, strengths, limitations and data quality provided with example of use.". *Int. J. Risk Safety Med.,* 1997, *10,* 1.

[7] Kjaergaard J., Clemmensen I.H., Storm H.H.: "Validity and completeness of registration of surgically treated malignant gynaecological diseases in the Danish National Hospital Registry". *J. Epidemiol. Biostat.,* 2001, *6,* 387.

[8] Dueholm M., Rokkones E., Lofgren M., Harkki P., Arason G.: "Nordic gynecologists' opinion on quality assessment registers". *Acta Obstet. Gynecol. Scand.,* 2004, *83,* 563.

[9] Gunnarsson U.: "Quality assurance in surgical oncology. Colorectal cancer as an example." *Eur. J. Surg. Oncol.,* 2003, *29,* 89.

[10] Gunnarsson U., Seligsohn E., Jestin P., Pahlman L.: "Registration and validity of surgical complications in colorectal cancer surgery". *Br. J. Surg.,* 2003, *90,* 454.

[11] Andersen T.F., Madsen M., Jorgensen J., Mellemkjoer L., Olsen J.H.: "The Danish National Hospital Register. A valuable source of data for modern health sciences". *Dan. Med. Bull.,* 1999, *46,* 263.

[12] Szklo M., Nieto F.J.: "Sensitivity, specificity, and percent agreement". In: J. Colilla (ed.), Epidemiology Beyond the Basics, Gaithersburg, Aspen Publishers, Inc., 2000, 452.

[13] Kjaergaard J., Clemmensen I.H., Thomsen B.L., Storm H.H.: "Validity of diagnoses and of operations for nonmalignant gynecological conditions in the Danish National Hospital Registry". *J. Clin. Epidemiol.,* 2002, *55*, 137.

[14] Mukherjee A.K., Leck I., Langley F.A., 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.

[15] Tingulstad S., Halvorsen T., Norstein J., Hagen B., Skjeldestad F.E.: "Completeness and accuracy of registration of ovarian cancer in the cancer registry of Norway". *Int. J. Cancer,* 2002, *98,* 907.

[16] MacDonald N., Sibley K., Rosenthal A., Menon U., Jayarajah A., Oram D. *et al.*: "A Comparison of national cancer registry and direct follow-up in the ascertainment of ovarian cancer". *Br. J. Cancer,* 1999, *80,* 1826.

[17] Storm H.H.: "Completeness of cancer registration in Denmark 1943-1966 and efficacy of record linkage procedures". *Int. J. Epidemiol.,* 1988, *17,* 44.

[18] Jensen A.R., Overgaard J., Storm H.H.: "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.

[19] Brenner H., Savitz D.A.: "The effects of sensitivity and specificity of case selection on validity, sample size, precision, and power in hospital-based case-control studies". *Am. J. Epidemiol.,* 1990, *132,* 181.

Address reprint requests to:
M.S. TETSCHE, M.D.
Department of Clinical Epidemiology
Aalborg Hospital
Aarhus University Hospital
Stengade 10,
DK 9000 Aalborg (Denmark)