

Valid and complete data on endometrial cancer in the Danish Gynaecological Cancer Database

Caroline S. Juhl^{1,2}, Estrid S. Hansen³, Claus K. Høgdall⁴ & Gitte Ørtoft²

ABSTRACT

INTRODUCTION: It is a comparative register study designed for data validation of surgery, pathology and recurrence for endometrial cancer in the Danish Gynaecological Cancer Database (DGCD) in the 2005-2009 period. The main outcomes were completeness of the data registered in the DGCD, agreement concerning data reported and comparability between the DGCD and a definite reference.

MATERIAL AND METHODS: DGCD data on women with endometrial cancer or adenomatous hyperplasia supplemented with patient charts for data on recurrence were retrieved and compared with a definite reference (the pathology report and clinical journals).

RESULTS: The completeness of data on pathology and surgery reported to the DGCD was 97.3%. The comparability between the DGCD and the definite reference was 94.4%. The agreement for the reported data in the DGCD was 88.3%. For recurrence, the comparability was 94.5% and the agreement was 71.6%. Completeness could not be determined due to the design of the database, where recurrence is composed of optional variables only.

CONCLUSION: The data on endometrial cancer registered in the DGCD regarding surgery and pathology are valid and complete, and they provide a solid base for research. Due to the relatively infrequent incidence of recurrences, and the fact that these are rarely entered into the database when they do occur, agreement concerning recurrence is low. Based on this study, the DGCD cannot alone provide information on recurrence that will give a reliable foundation for research.

FUNDING: Funding was provided by the Health Research Fund of the Region of Central Jutland and the Department of Gynaecology and Obstetrics, Aarhus University Hospital.

TRIAL REGISTRATION: not relevant.

Endometrial cancer is a common gynaecological cancer in Denmark with a life-time risk of 2%, accounting for 3.7% of all cancer cases in women [1]. The incidence was rising in Denmark until 1980, and then stabilised at approx. 700 new annual cases (13:100,000). The incidence is higher in Norway (16:100,000), Finland and Sweden (approx. 14:100,000) than in Denmark (NORDCAN/DK, 9.10.2012). Endometrial cancer in young women is rare, and the incidence increases from the age of 45 years. In Denmark, 67.4% of the cases are diagnosed in stage I

(and adenomatous hyperplasia), 13.1% in stage II, 12.5% in stage III and 2.2% in stage IV. The overall five-year survival is 75.1%, varying across the stages [1].

To secure equal quality of treatment of gynaecological cancers in Denmark, the Danish Gynaecological Cancer Database (DGCD) was established. It monitors and follows the changes in survival, recurrence and treatment. The DGCD is also designed for clinical and scientific research [2]. During the past decade, major changes in both surgical and adjuvant treatment for endometrial cancer have been implemented in Denmark. Surgery has become more extensive, and laparoscopic and robotic surgery are being used in many hospitals. Post-operative radiation is now rarely used, and chemotherapy is being tested [3].

Denmark is an ideal setting for epidemiological and register studies as databases are used extensively, including the Civil Person Registry (CPR) which monitors people from birth to death (or emigration), providing an opportunity to link different registers [4]. It is essential that a database contains reliable data, and validation is therefore of crucial importance. The aim of this study was to validate data on endometrial cancer registered in the DGCD for the 2005-2009 period.

MATERIAL AND METHODS

Study population

Data from women with endometrial cancer or adenomatous hyperplasia registered in the DGCD (International Classification of Diseases (ICD)-10 codes D070, N85.1 and D707, C54-C55) were included. There were no exclusion criteria if the above diagnosis was correct.

The data were retrieved from the DGCD on 13 September 2010, covering the period from 2005 to 2009. The dataset comprised a total of 3,388 registered patients with endometrial cancer or adenomatous hyperplasia, among whom 12.5% (422 patients) were randomly selected. The selection was stratified for disease stage and hospital volume. One patient was excluded because she did not meet the inclusion criteria, which left 421 cases for final validation. When the annual report from the DGCD was published (9 December, 2011), an additional 339 patients (10%) were added to the dataset after central validation by the DGCD. These patients were not included in the present study.

ORIGINAL ARTICLE

- 1) Department of Obstetrics and Gynaecology, Randers Region Hospital
- 2) Department of Obstetrics and Gynaecology, Aarhus University Hospital
- 3) Department of Pathology, Aarhus University Hospital
- 4) Department of Gynaecology Clinic, The Juliane Marie Centre, Rigshospitalet

Dan Med J
2014;61(6):A4864

 TABLE 1

Completeness, comparability and agreement on Danish Gynaecological Cancer Database variables reported by surgeon and pathologist to the Danish Gynaecological Cancer Database^a. The values are % (n/N).

DGCD variables	Completeness of data in the pathology report	Completeness of data/DGCD variables on the patients reported in the DGCD	Total comparability between the DGCD and the definite reference for all data incl. missing data	Agreement on reported data between the DGCD and definite reference
<i>Surgeon</i>				
Answer required:				
Atypia/carcinoma	96.2 (405/421)	98.8 (416/421)	92.9 (391/421)	93.3 (388/416)
Hysterectomy	100.0 (421/421)	94.8 (399/421)	93.6 (394/421)	98.7 (394/399)
Oophorectomy sin.	100.0 (421/421)	98.3 (414/421)	92.6 (390/421)	94.2 (390/414)
Oophorectomy dxt.	100.0 (421/421)	97.6 (411/421)	91.7 (386/421)	93.9 (386/411)
Salpingectomy dxt.	100.0 (421/421)	98.1 (413/421)	91.7 (386/421)	93.5 (386/413)
Salpingectomy sin.	100.0 (421/421)	97.9 (412/421)	91.9 (387/421)	93.9 (387/412)
Cytology	100.0 (421/421)	97.1 (409/421)	92.6 (390/421)	95.4 (390/409)
Omentectomy	100.0 (421/421)	95.7 (403/421)	91.9 (387/421)	96.0 (387/403)
Appendectomy	100.0 (421/421)	89.3 (376/421)	84.8 (357/421)	94.9 (357/376)
Lymph node ^b	100.0 (421/421)	96.9 (408/421)	95.7 (403/421)	98.8 (403/408)
Average on DGCD variables, required	99.6 (996.2/1,000)	96.5 (964.5/1,000)	91.9 (919.4/10)	95.3 (952.6/1,000)
Answer optional:				
Paraaortic lymph nodes	100.0 (421/421)	–	99.0 (417/421)	77.8 (7/9)
Iliac lymph nodes	100.0 (421/421)	–	99.0 (417/421)	94.7 (72/76)
Inguinal lymph nodes	100.0 (421/421)	–	99.3 (418/421)	0.0 (0/3)
Average on DGCD variables, optional	100.0 (300/300)	–	99.1 (297.3/3)	57.5 (172.5/3)
Average on DGCD variables, optional + required	99.7 (1,296.2/1,300)	–	93.6 (1,216.7/1,300)	86.5 (1,125.1/1,300)
<i>Pathologist</i>				
Answer required:				
Stage	99.3 (418/421)	99.8 (420/421)	93.3 (393/421)	93.6 (393/420)
Dominant histology	100.0 (384/384)	100.0 (384/384)	91.4 (351/384)	91.4 (351/384)
Grade of tumour differentiation	100.0 (384/384)	99.5 (382/384)	92.7 (356/384)	93.2 (356/382)
Cytology	100.0 (384/384)	100.0 (384/384)	94.5 (363/384)	94.5 (363/384)
Cervical involvement	99.2 (381/384)	97.1 (373/384)	94.8 (364/384)	97.3 (363/373)
Average on DGCD variables, required	99.7 (498.5/500)	99.1 (495.4/500)	93.3 (466.7/500)	94.0 (470/500)
Answer optional:				
Lymph node removed	100.0 (384/384)	–	90.1 (348/384)	91.5 (313/342)
Lymph metastasis, total	100.0 (384/384)	–	97.7 (375/384)	89.8 (307/342)
Pelvic lymph nodes	100.0 (384/384)	–	99.5 (382/384)	90.9 (10/11)
Paraaortic lymph nodes	100.0 (384/384)	–	100.0 (384/384)	100.0 (4/4)
Uterus not received	100.0 (384/384)	–	99.2 (381/384)	93.3 (14/15)
Myometrial invasion	100.0 (384/384)	–	92.4 (355/384)	95.2 (354/372)
Parametrial invasion	100.0 (384/384)	–	98.4 (378/384)	100.0 (13/13)
Adenomyosis	100.0 (384/384)	–	97.1 (373/384)	84.6 (11/13)
Vaginal metastasis	100.0 (384/384)	–	97.9 (376/384)	50.0 (2/4)
Parametrium, metast.	100.0 (384/384)	–	99.0 (380/384)	90.9 (20/22)
Tuba involvement	100.0 (384/384)	–	100.0 (384/384)	100.0 (6/6)
Ovary, metastasis	100.0 (384/384)	–	99.2 (381/384)	93.8 (15/16)
Omentum, metastasis	100.0 (384/384)	–	99.7 (383/384)	100.0 (4/4)
Other, metastasis	100.0 (384/384)	–	98.4 (378/384)	50.0 (4/8)
Average on DGCD variables, optional	100.0 (1,400/1,400)	–	97.6 (1,368.6/1,400)	87.9 (1,230.0/1,400)
Average on DGCD variables, optional + required	99.9 (1,898.5/1,900)	–	96.6 (1,835.3/1,900)	89.5 (1,700/1,900)
<i>Surgeon and pathologist (in total)</i>				
Average on DGCD variables, required	99.6 (1,494.7/1,500)	97.3 (1,459.9/1,500)	92.4 (1,386.1/1,500)	94.8 (1,422.6/1,500)
Average on DGCD variables, optional	100.0 (1,700/1,700)	–	98.0 (1,665.9/1,700)	82.5 (1,402.5/1,700)
Average on DGCD variables, optional + required	99.8 (3,194.7/3,200)	–	95.4 (3,052/3,200)	88.3 (2,825.1/32)

DGCD = Danish Gynaecological Cancer Database.

a) As a definite reference, the pathology report from the Danish Pathology Database was used. When patients with adenomatous hyperplasia are excluded, the study population is reduced from 421 to 384. A required answer is when an answer must be given, including a negative answers. Optional answers are only given if positive.

b) "Lymph nodes removed" is a composed value created by use of the sub-variables: lymph nodes removed paraaortic, iliac, inguinal and no removal of lymph nodes.



TABLE 2

Comparability and agreement on variables for recurrence reported to the Danish Gynaecological Cancer Database^a.

Recurrences	DGCD variables			correct variable in the DGCD out of all possible variables founding the definite reference, % (n/N)	total comparability between the definite reference and the DGCD for all data incl. missing data, % (n/N)	agreement on reported data between the definite reference and the DGCD, % (n/N)
	patients with recurrence reported to the DGCD, n	patients with recurrence found in the pathology report, n	patients with recurrence found in patient charts, n			
<i>Answer optional</i>						
Recurrences	20	45	52	34.6 (18/52)	91.4 (385/421)	90.0 (18/20)
Vaginal recurrences	11	25	25	32.0 (8/25)	95.2 (401/421)	72.7 (8/11)
Central pelvic recurrences	3	2	7	14.3 (1/7)	98.1 (413/421)	33.3 (1/3)
Pelvic recurrences	2	2	5	20.0 (1/5)	98.8 (416/421)	50.0 (1/2)
Other recurrences	5	20	37	13.5 (5/37)	92.4 (389/421)	100.0 (5/5)
Biopsy-verified	12	45	44	22.2 (10/45)	91.2 (384/421)	83.3 (10/12)
Average on DGCD variables	–	–	–	–	94.5 (567.1/600)	71.6 (429.3/600)

DGCD = Danish Gynaecological Cancer Database.

a) Definite reference is based on the pathology report from the pathology database and the patient charts and 23 entries have missing data on all variables. This occurs when no gynaecological or oncological data can be found in the journal. Optional answers are only reported if positive.

Procedure

The pathology reports were retrieved from the national pathology database, entered into a new database (by a single person to secure homogeneity) and then compared with DGCD data. Disagreements were double-checked. The pathology report was chosen as a definite reference. Patient charts were added as a reference for recurrences to include those not pathologically verified. The values for some of the optional and all the required DGCD variables in **Table 1** are averages calculated from “sub-variables”/the different response options given. A response option is to choose between yes or no, or between grade 1, 2 or 3 for tumour differentiation. Calculations were done for all “sub-variables”, and an average was calculated for the main variable to which they belong (Table 1).

Data from the DGCD were validated for completeness and agreeability. Agreeability was validated in two different settings; only data reported, to see how valid actual data in DGCD are (agreement), and *all* data, both reported and not reported, including missing data, to see how equivalent the two databases are on all possible values (total comparability).

Missing data were registered as correct if data were missing in both databases. For optional variables, a blank space (both missing and not-positive values) was registered as correct in the DGCD if comparable to the definite reference. An earlier study evaluated the completeness of entries into the DGCD, so this was not re-examined. Completeness in the present study was defined as completeness of data in the DGCD on patients already registered in the DGCD. The variables assessed are given in Table 1 and **Table 2**. Two DGCD variables

(“tumour expansion” and “simultaneous ovarian cancer”) were not validated due to low completeness in the definite reference.

The Danish Gynaecological Cancer Database

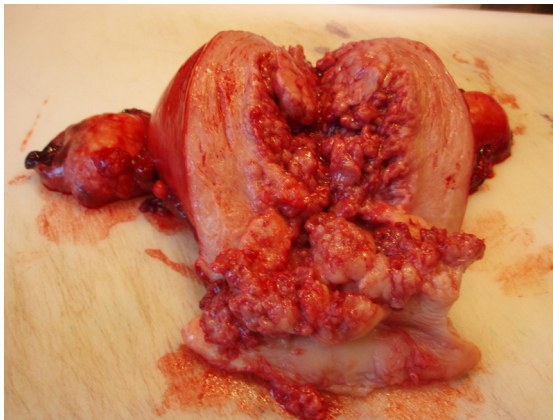
The DGCD was made operational on 1 January 2005. It is a multidisciplinary, nationwide database containing reports from gynaecologists, pathologists, oncologists and nurses. Clinicians can access the database on-line [5]. Reporting to the database is mandatory by law, and the database is required to include at least 90% of the relevant population. The DGCD has an actual completeness of 94.2% [2]. All entries are by CPR number and contain information on general health, medical history, surgery, pathology, complications, recurrences and death. The database was designed to prevent most conflicting information.

Most variables are required, meaning that an answer has to be reported, even if it is unknown, before the person entering the data can proceed. A few required variables can be forced, but against DGCD recommendations. A few variables are optional. Hence, an answer is entered or the space is left blank. A central validation of completeness is performed annually by the DGCD, but the reliability of data entered into the database is not validated by the DGCD.

The pathology database

The pathology database, established in 1999, is a Danish nationwide database containing the results of all cytology and histopathology examinations performed in Denmark. The database contains CPR numbers, places of admittance, diagnoses and pathological descriptions.

Endometrial cancer.
Source: Department
of Pathology, Aarhus
University Hospital.



Statistical analysis

The data retrieved from the pathology reports were entered into Epi-Data and then transferred to the STATA (Stata Statistical Software, Release 11.1 Collage Station, USA; 2009).

Trial registration: not relevant.

RESULTS

For final validation, 421 cases were chosen, including 37 with adenomatous hyperplasia. When the preoperative diagnosis is adenomatous hyperplasia, only the variables relevant for adenomatous hyperplasia open up in the database. The study population for the remaining variables is therefore 384 only.

The total average for completeness for both pathological and surgical variables is 97.3%. For pathological variables alone, the average completeness is 99.1% and for the surgical cases only, it is 96.5% (Table 1). The total comparability between the DGCD and the definite reference is 95.4% (Table 1). For surgical variables alone, the comparability was 93.6%; and when we discriminated between required and optional variables, it was 91.9% and 99.1%, respectively. For pathological variables alone, the comparability was 96.6%; and when we discriminated between required and optional variables, it was 93.3% and 97.6%, respectively. The total agreement in the DGCD is 88.3%. Discrimination between required and optional variables yields a total agreement for required variables of 94.8%, whereas it is 82.5% for optional value (Table 1). The table for recurrences has optional variables only (Table 2). The total comparability is 94.5%. The average for total agreement on reported data is 71.6%. Optional variables cannot be validated for completeness due to the uncertain value of a blank space.

DISCUSSION

The aim of the study was to validate the DGCD data for

women with endometrial cancer or adenomatous hyperplasia. The DGCD was generally found to be a reliable database. Data on pathology and surgery, completeness and comparability were excellent. For total agreement, the quality was somewhat lower, but still good. For recurrence, the comparability was very good, but agreement not as good. The results differed depending on the voluntariness. The required variables were valid and homogeneous, but the optional variables varied due mainly to the variation in the number of entries. Variables with few entries reported are sensitive and therefore inconclusive. This is often seen for optional variables. For optional, and especially for the seldom variables, total comparability is of importance. With few positive entries, the reliability of the blank space becomes of interest. The total comparability in the DGCD is found to be very reliable. Only one variable is lower than 90% (appendectomy, 84.8%).

The most valid data in the DGCD are those that apply to required variables. There is no difference in validity of the required data whether it concerns the surgical or the pathological ones. No required variable, except for appendectomy, has a validation value below 91.4%. Three of the required surgical variables can be forced without answering (appendectomy, cytology and omentectomy). These three variables are among the five variables which have the lowest completeness. The variable "appendectomy" has the lowest validity of the three (Table 1). One "sub-variable" is "earlier appendectomy". The surgeon has most likely not been involved with the "earlier appendectomy" (the sub-variable). It may be reported as not removed even if it was, in fact, removed earlier, or the space is forced without answering because the reporter is uncertain about the answer. The "sub-variable" "earlier omentectomy" has the same problem, but the risk of a wrong answer is greater for appendectomy as it is a more common event than omentectomy. The variable "hysterectomy" is required and cannot be forced. Its completeness is lower (94.8%) than that of other required variables. This is due to the "sub-variable" "no hysterectomy". For "no hysterectomy" to be reported, the surgeon has to log on to the DGCD and compose a surgery-entry saying no hysterectomy. For the study population, this was done twice, both incorrectly.

The comparability of optional variables is high. All variables for optional data reported with more than ten entries have good agreement. The lowest agreement appears when few entries are reported. The voluntariness could explain the low number of entries, but other factors may also be of importance. The surgeons have inguinal lymph nodes with 0% agreement. It is a seldom place of metastasis for endometrial cancer. In the DGCD, the "inguinal lymph node" registration button is situated

besides two more common places for lymph node metastasis, and they can be chosen accidentally. We found that the risk of a wrong entry was greater than the frequency of inguinal lymph node metastasis. The inguinal lymph node metastasis is not included in endometrial cancer staging, which questions the relevance of this button.

The definite reference for recurrence consists of the pathology report and the patient charts. The variables for recurrences are all optional in the DGCD with a high total comparability, lower average agreement and a small entry number. The low agreement is due to the small number of entries. The small number of entries could be due to there being no incidence of recurrence, but it is more likely that most Danish oncological departments do not report to the DGCD, and many patients with recurrences are seen at these departments. Treatment can be multidisciplinary, and the responsibility for reporting to the DGCD therefore becomes uncertain. The clinical journals revealed 52 entries with recurrence. Only 20 were reported to the DGCD, and 10.0% of these were incorrect. In total, only 34.6% of all possible entries for "recurrences" were reported correctly (Table 2). One DGCD variable, "other recurrences", demonstrated 100% agreement, as all reported cases were correct, but 86.5% of the actual "other recurrences" (due to the definite reference) were not reported to the DGCD, and hence only 13.5% of all possible entries were reported correctly. A registration model in which the pathologist reports the recurrences may yield better agreement.

An earlier study validated the strength of agreement of ovarian cancer quality indicators and registration in the DGCD. The study found that data linked to the surgical and pathological procedures had a great strength of agreement except for "grade" and "complications" [2]. The present study examined different aspects and variables, and a one-to-one comparison cannot be made. However, with the high validity found in both studies, except for specific variables and recurrences, it seems likely that this is the case for all DGCD data, including cervix cancer data.

The present study has a limitation regarding unknown data. There are missing data for both definite references, but more for the patient charts (5.5%). The patient charts used are electronic and not updated on all hospitals. As the patients were stratified with respect to hospitals when selected, this should not create bias. By the time of data retrieval, the DGCD had not finished the central validation for the study period in question. When finished (December 2010), 10% new entries were added. These 10% represent the usual follow-up and do not to introduce bias. The composition of DGCD with forcible and optional variables also creates uncertain data.

Some of the study population has a short follow-up

time. This results in a lower number of entries for recurrence than would have been the case with a longer follow-up time. It does not change the composition of the population, but a higher number of entries would give less sensitive data and perhaps a better agreement.

The DGCD is currently updating in order to reflect the changes in treatments, repair bugs and increase completeness and validity in accordance with studies like the present. A new and improved version of the DGCD was launched in 2013 where optional variables had been made mandatory and a new logistics introduced for registration of recurrences.

CONCLUSION

The surgery and pathology data on endometrial cancer registered in the DGCD are extraordinarily valid and complete. The data provide a solid base for research. The agreement on recurrence is low due to the relatively infrequent incidence which, when occurring, is only rarely entered into the database. Based on this study, the DGCD alone, cannot provide information on recurrence that yields a reliable foundation for research. The new version of the DGCD is an attempt to solve the presented shortcomings.

CORRESPONDENCE: *Caroline Sollberger Juhl, c/o Gitte Ørtoft, Brendstrupgårdsvej 100, 8200 Aarhus N, Denmark. E-mail: caroj77@yahoo.com*

ACCEPTED: 11 April 2014

CONFLICTS OF INTEREST: Disclosure forms provided by the authors are available with the full text of this article at www.danmedj.dk.

LITERATURE

1. Høgdall CK, Taaning L, Nielsen MLS. The Danish Gynaecologic Cancer Database – a nationwide clinical database for ovarian cancer, endometrial cancer and cervical cancer, year 2011. www.dgdc.dk/images/DGCD%20rsrapport%202011.pdf (Dec 2012).
2. Petri A, Kjaer L, Christensen SK et al. Validation of epithelial ovarian cancer and fallopian tube cancer and ovarian borderline tumor data in the Danish Gynecological Cancer Database. *Acta Obstet Gynecol Scand* 2009;88:536-42.
3. Amant F, Mirza MR, Creutzberg CL. Cancer of the corpus uteri. *Int J Gynaecol Obstet* 2012;119:110-7.
4. Frank L. Epidemiology – when an entire country is a cohort. *Science* 2000;287:2398-9.
5. Petri AL, Høgdall C, Lidsgaard Ø. Registration of primary ovarian cancer in Denmark. *Ugeskr Læger* 2009;171:408-11.