Data completeness can be optimized in clinical databases

Peer Wille-Jørgensen

ABSTRACT

INTRODUCTION: The aim of the present paper was to analyse the incidence of and reasons for data incompleteness in reporting of colorectal cancer from a single department to a national database.

MATERIAL AND METHODS: A list of errors comparing the database with the National Patient Registry is drawn up monthly and has been used for data completeness analysis for several years as part of the department's daily routine. In May 2009, this system was revised and we identified a number of new errors which we previously thought had been corrected. The nature of these errors has been analysed in detail.

RESULTS: According to the National Patients Registry, a total of 1,530 patients were treated for a first-time colorectal cancer. In 60 cases (3.9%) patients were missing in the database. Erroneous registration of the diagnosis in the National Patient Registry on the part of the department was the primary reason why data were missing. This was followed by clerical errors at the department. All mistakes were explainable, but changes in the registry structure prevented some of the mistakes discovered from being corrected retrospectively.

CONCLUSION: The number of missing patients in the database could be minimized by a meticulous effort at controlling data and comparing them with data in national registries.

The use of clinical databases for quality assurance and epidemiological research has been growing steadily for more than 20 years. The use of indicators for benchmarking and for guiding the public in their choice of hospital for treatment of a specific disease is now common in many specialties. However, the effect of the use of such indicators on treatment quality can be discussed and their use harbours potential dangers [1]. The prerequisites for using databases for these purposes include a high level of data completeness and a high validity. Ensuring completeness and validation of data is a major part of the work of managing a database [2-4]. In Denmark about 20 databases covering different areas on a nationwide basis are supported financially by the Health Authorities, among others the Danish Colorectal Cancer Database (DCCG) [5]. This database was established in

2001, and during its fist years, data were reported on paper forms and the validity of data was only checked sporadically which meant that only obvious mistakes were found. As from 2005, the data were reported directly and electronically with quality control of the individual data performed during reporting and it was made possible to draw lists revealing incompleteness from the individual department as database entries were regularly controlled by comparing their data with information from the Danish National Patient Registry. As from May 2009, the procedure for reporting on missing patients and reported errors was changed so that it became more accurate and more sensitive. The aim of this study was to identify reasons for missing patients and for registration errors by tracing consecutive reporting from a high-volume department.

MATERIAL AND METHODS

From 1 May 2001 to 31 December 2008, The Department of Surgery K at Bispebjerg Hospital, Copenhagen handled 1,530 patients with colorectal cancer. The reporting system was organized locally as follows:

The printed reporting forms which was the basic instrument used for the electronic reporting consisted of two parts: a green form for patient-related information (e.g. symptoms, comorbidity and life-style risk factors) and a yellow form covering treatment-related information (e.g. type of surgery, tumour stage and complications). The operating surgeon or attending physician was responsible for filling in the forms and for reporting the

Working with files.

correct diagnosis code to the National Patient Registry. Five to six weeks after surgery or after the decision not to operate, the forms were collected and data were validated against the patient files by one of two persons during the whole period. The data were then entered into the database by the same two persons. Twice a year, a list detailing completeness was drawn, and the database was updated with any missing files. The last of these updates was done in early 2009 in order by 1 May 2009 to have »full house«, i.e. data coverage up to 1 January 2009. Except for one file from 2004 which was known to be lost, we believe that full data coverage was achieved through this process.

In May 2009 the more advanced completeness lists were used and they revealed missing or wrong reporting in 60 patients (3.9%) in the investigated cohort. All these 60 records had been filled in before the control system was changed. The patient files were found and the reasons for observed errors were analysed in each case. These reasons are reported below. For comparison with other departments, we checked the data completeness of all 28 departments delivering data to the DCCG-database as found in the 2008 annual report from the DCCG.

RESULTS

The various reasons for incorrect or lacking registration are listed in **Table 1**. The most common errors were due to wrong or missing coding into the National Patient Registry (benign diseases, cancers other than colorectal and missing coding because the patient never reached the department, but remained in a medical ward) (n = 22). Another common error was missing registration of patients in the database due to clerical error at the

TABLE 1

Reasons for missing or wrong registrations in 60 patients out of 1,530 patients handled for colorectal cancer from 1 May 2001 to 31 December 2008.

Reason for incorrect or lacking registration	n
Wrong coding at department	15
Not in the National Patient Registry	1
Not adenocarcinoma	2
Metachronous cancers	9
Operated abroad or at private hospitals	3
Operated at another public hospital	2
Missing data (e.g. day of operation)	2
Only submitted to medical wards	7
Not in the local patient registry	2
Immigrated before operation	1
Lack of registration in database	11
Files missing or incomplete in electronic archive	5
Total	60

department because the patient files had never reached the responsible surgeons (n = 11). Another mistake was entry of metachronous cancers incorrectly reported to the database (n = 7). The rest of the errors had arisen for various reasons as stated in the Table. It is remarkable that five patient files were missing or incomplete in the electronic archive of scanned patient files (the paper version had been destroyed).

The patient registration completeness in the database for the whole of Denmark varied in 2008 from 22% to 99% between the individual departments with an overall completeness of 92%.

DISCUSSION

Data completeness and validity are essential for the clinical databases to be able to serve their purposes: quality assurance, benchmarking and epidemiological research. Owing to the Scandinavian system in which a unique identification number is assigned to each citizen, it is rather easy to survey the completeness of entrance into the database by comparing entries into the patient administrative systems with entries into the database. The Danish National Patient Registry has existed for more than 20 years and is generally considered to have a high completeness regarding patients treated in the public health care system [2, 6-8]. In countries outside Scandinavia, obtaining an acceptable data completeness has been reported to be more difficult [9, 10].

However, the validity of data in the individual databases cannot be checked in detail by electronic control. Some types of data validity can be obtained by means of internal, logical electronic feedbacks, thus preventing the reporting of conflicting data (e.g. performing hysterectomy on a male), but many details may be reported wrongly. A manual comparison of two versions reported by two different persons (surgical theatre scrub nurses and ward secretaries) to the same surgery registry showed an 11.2% discrepancy in registrations [11]. The data quality of the DCCG database has been checked manually since before electronic reporting began and it has been found to be satisfactory with Kappa values between 0.54 and 0.94 for chosen variables. The only outlier was the American Society of Anesthesiologists' (ASA) classification which showed a Kappa value of only 0.09 [12]. In another database on colorectal cancer surgery, the same validity could not be obtained [3].

The data completeness reported in this paper must be considered satisfactory, but it reflects a continual and intense work of maintaining the database and repeatedly looking for instances of missing reporting. The data completeness is comparable with that seen in other Danish databases [13-16]. The results for this high-volume department cannot be directly compared with a prevalence investigation from departments from other

hospitals in the rest of the country because in this list of missing or wrongly reported patients, we only checked entries up to the year 2008. We chose to do so because we know that entries into the database are always lacking a couple of months behind entries in the National Patient Registry, which we thought would reflect a bias in this paper. In June 2009, 47 out of 108 either totally missing or wrongly registered patients were from this department. More interestingly, there is much diversity in the number of missing patients throughout the country. This must reflect different local traditions and differences in how much effort is put into obtaining data completeness at the local level. Another explanation could be that many large departments have been merged over the last couple of years. The consequences of this diversity could be that the results of studies performed on the basis of data from these departments are less reliable than expected even if the registries enjoy high overall completeness (92%). Diversity is also a problem where databases are used to benchmark departments with a low reporting frequency. One could suspect that the missing files stem from the most complicated patients, which would bias the interpretation of the results.

Most physicians support the use of clinical databases and wish to participate by reporting their data, but many become more reluctant if use of the database involves a cost or demands extra work [17]. Such attitudes could also explain part of the great diversity in data completeness among the Danish departments.

Most of the missing or erroneous reporting could have been avoided by a more meticulous coding practice and a more intense follow-up on regularly drawn lists of completeness, but some of the errors cannot be explained. We have tried to correct all the errors, but due to a change in the patient administrative systems, it is not possible after 1 Jan 2010 to change entries made before 14 March 2009. We must accept that some errors will therefore remain in the database.

The ideal database should retrieve its data directly from electronic patient files in which coding is incorporated automatically. This is technically possible, but the use in practice has yet to be proven. Until this works in practice, the only way to obtain a high data completeness is to prioritize registration at the individual departments. The number of data per patient (the DCCG-database for example contains 60 data groups, many with several data entry options) also sets a limit for automatic systems as many data (e.g. complications and surgical details) demand individual judgement by a professional, medically trained person. Reducing the amount of data in order to make automatic systems workable would limit the possibilities of using the results for epidemiological studies. This would make the database an instrument only for studies of case volume and, to some extent, outcomes.

In conclusion, it is possible to obtain a high level of data completeness in a clinical database at the level of the individual department. Errors can often be corrected. The number of errors differs much between individual departments.

CORRESPONDENCE: *Peer Wille-Jørgensen,* Department of Surgery K, Bispebjerg Hospital, 2400 Copenhagen NV. Denmark.
E-mail: pwil0002@bbh.regionh.dk

LITERATURE

1. Endahl LA, Utzon J. [Will publication of quality indicators in the health service improve the quality? International experiences and Danish perspectives]. Ugeskr Læger 2002;164:4380-4.
2. Tingulstad S, Halvorsen T, Norstein J et al. Completeness and accuracy of registration of ovarian cancer in the cancer registry of Norway. Int J Cancer 2002;98:907-11.
3. Gunnarsson U, Seligsohn E, Jestin P et al. Registration and validity of surgical complications in colorectal cancer surgery. Br J Surg 2003;90:454-9.
4. Wille-Jorgensen PA, Sorensen LT, Roodpashti AM et al. [Difficulties with implementation and maintenance of a clinical database]. Ugeskr Læger 1999;161:6359-62.
5. Harling H, Nickelsen T. [The Danish Colorectal Cancer Database]. Ugeskr Læger 2005;167:4187-9.
6. Lidegaard O, Vestergaard CH, Hammerum MS. [Quality monitoring based on data from the Danish National Patient Registry]. Ugeskr Læger 2009;171:412-5.
7. Lidegaard O, Hammerum MS. [The National Patient Registry as a tool for continuous production and quality control]. Ugeskr Læger 2002;164:4420-3.
8. Nickelsen TN. [Data validity and coverage in the Danish National Health Registry. A literature review]. Ugeskr Læger 2001;164:33-7.
9. Mukherjee AK, Leck I, Langley FA et al. The completeness and accuracy of health authority and cancer registry records according to a study of ovarian neoplasms. Public Health 1991;105:69-78.
10. Kashner TM. Agreement between administrative files and written medical records: a case of the Department of Veterans Affairs. Med Care 1998;36:1324-36.
11. Wille-Jorgensen PA, Meisner S. [The validity of data in registration of operations. A quality analysis]. Ugeskr Læger 1997;159:7328-30.
12. Nickelsen TN, Harling H, Kronborg O et al. [The completeness and quality of the Danish Colorectal Cancer clinical database on colorectal cancer]. Ugeskr Læger 2004;166:3092-5.
13. Bardram L . Danish Cholecystectomy Register, Annual Report 2007:18-19.
14. Dreisler E, Schou L, Adamsen S. Completeness and accuracy of voluntary reporting to a national case registry of laparoscopic cholecystectomy. Int J Qual Health Care 2001;13:51-5.
15. Jensen AR, Storm HH, Moller S et al. Validity and representativity in the Danish Breast Cancer Cooperative Group – a study on protocol allocation and data validity from one county to a multi-centre database. Acta Oncol 2003;42:179-85.
16. Hansen CT, Moller C, Daugbjerg S et al. Establishment of a national Danish hysterectomy database: preliminary report on the first 13,425 hysterectomies. Acta Obstet Gynecol Scand 2008;87:546-57.
17. Dueholm M, Rokkones E, Lofgren M et al. Nordic gynecologists' opinion on quality assessment registers. Acta Obstet Gynecol Scand 2004;83:563-9.