Validation of a register-based algorithm for recurrence in rectal cancer

Emilie Palmgren Colov[1], Tina Fransgaard[2], Mads Klein[3] & Ismail Gögenur[2]

ABSTRACT

INTRODUCTION: Colorectal cancer is the third most common type of cancer worldwide. Recurrence is an important end-point in colorectal cancer research. This study aimed to validate a previously developed algorithm to determine recurrence after surgery for rectal cancer.
METHODS: Information about a cohort of 500 patients with rectal cancer was retrieved from the Danish Civil Registration System, the Danish National Patient Register and the Danish National Pathology Register. Patients with an ICD-10 code for metastatic disease, chemotherapy or SNOMED code suggesting metastasis were identified as having recurrence. In a previous study, medical records of the same cohort had been reviewed to identify recurrence. Recurrence identified by the algorithm was compared with the previously retrieved information from the medical records using Kappa statistics. The sensitivity and specificity of the algorithm were determined.
RESULTS: Of the 500 patients, 393 were included in the validation analysis. Kappa statistics showed good concordance with a Kappa value (95% confidence interval) of 0.80 (0.72-0.88). The sensitivity and specificity of the algorithm were 88% (77-95%) and 96% (93-98%), respectively.
CONCLUSIONS: The study showed good concordance between the algorithm and medical record information. The algorithm makes it possible to perform large register-based studies of recurrence and disease-free survival in colorectal cancer patients without the need for evaluation of medical records.
FUNDING: none.
TRIAL REGISTRATION: not relevant.

In most cases, surgery is the preferred treatment for colorectal cancer, but 30-40% of potentially curable colorectal cancer patients relapse [1]. Death and disease-free survival are often used as important outcomes in cancer research. Advances in the treatment of recurrence after colorectal cancer have translated into improved survival in patients with recurrent colorectal cancer [2, 3]. This makes recurrence, and not only death, an important oncological endpoint within the field of cancer research.

Denmark has a long-standing tradition for nation-wide registers. Unfortunately, recurrence is not currently recorded systematically in these registers. In order to determine the incidence of recurrence after surgery for colorectal cancer, an algorithm based on the Danish registers was developed [4]. Until now, studies focusing on recurrence have been limited to determining if a patient experienced recurrence by reviewing the medical records. The cohort used for developing the algorithm consisted of colorectal cancer patients operated and registered in the Danish Colorectal Cancer Group's (DCCG) database between May 2001 and December 2011. The findings of the algorithm were validated in two different actively followed cohorts of colorectal cancer patients in the original study [4].

The aim of the present study was to validate the previously developed algorithm in a nationwide cohort of patients with rectal cancer, thereby assessing the external validity of the method.

METHODS

In Denmark, all persons with a permanent residence are registered in the Danish Civil Registration System with a unique Central Person Register (CPR) number [5]. The CPR number is used whenever a person comes into contact with the Danish authorities, including Danish healthcare. This makes it possible to merge information from different registers.

In the present study, data were obtained from the following registers: the Danish Civil Registration System, the Danish National Patient Register (NPR) and the Danish Pathology Register (DPR).

The NPR contains information about contacts with the healthcare system. This includes information about diagnoses (using the International Classification of Diseases, tenth revision (ICD-10), treatments and procedures [6]. The DPR records information about all biological specimens. The specimens are described using the Danish version of the Systematized Nomenclature of Medicine (SNOMED) codes [7].

To validate the algorithm, a cohort consisting of patients with rectal cancer operated with curative intent between January 2009 and August 2012 was used. This cohort was derived from the DCCG's database which contain information on all patients undergoing surgery

for colorectal cancer in Denmark. The cohort was described in detail in a previous study [8] where the medical records were assessed in order to evaluate complications, recurrence and survival. In the previous study, recurrence was defined as clinical, radiological or pathological recurrence described by the clinicians in the medical record. In the present study, the incidence of recurrence found when using the algorithm was compared with the incidence of recurrence found when assessing the medical records.

According to the algorithm previously described by Lash et al [4], recurrence was defined as recurrence more than 180 days after colorectal cancer surgery. The algorithm was suited only for patients with colorectal cancer who had not had another primary cancer (defined as any cancer disease with the exception of non-melanoma skin cancer (ICD-10 C44)) prior to their diagnosis of colorectal cancer. Furthermore, patients who died or were diagnosed with a new primary tumour or metastasis within 180 days after surgery were excluded.

The patients were identified as having recurrence if at least one of the following four criteria was fulfilled:

1) A diagnosis code for metastases was found in the NPR 180 days or more after surgery without a diagnosis of a new primary cancer between surgery and the date of the metastasis code.
2) A cytostatic therapy code 180 days or more after surgery. The patient must not have been diagnosed with another primary cancer between colorectal cancer surgery and the date of the cytostatic therapy code.
3) The patients had SNOMED combinations describing metastasis or local recurrence for colorectal cancer recorded in the DPR 180 days or more after the colorectal cancer surgery without a new primary cancer in-between.
4) One of the specific codes for local recurrence of colorectal cancer.

A more detailed description of the algorithm can be found elsewhere [4].

In this study, no information was extracted from the Danish Cancer Register even though this formed part of the original algorithm. This was chosen because information has been transferred directly from the NPR to the Cancer Register since 2004. Therefore, the relevant data for the years included in this study were retrieved directly from the NPR.

Statistics

To examine the concordance between the two methods, Kappa statistics was used. The performance of the algorithm in terms of sensitivity and specificity was es-

timated from the contingency table using results from the medical records as the gold standard. Analyses were performed using SAS Proprietary Software Version 9.4. (SAS Institute Inc., Cary, NC USA). The study was approved by the Danish Data Protection Agency (2016-41-4745).

Trial registration: not relevant.

RESULTS

A total of 500 patients were available from the described cohort; 107 patients were excluded as they fulfilled the exclusion criteria. Thus, 393 patients were included in the validation analysis. For complete data about exclusions, see **Figure 1**. The follow-up time was between 18 months and five years depending on when the patient was operated. **Table 1** lists the number of patients with and without recurrence identified by the algorithm and by assessing the medical records.

Recurrence was identified in 53 patients by both the algorithm and the medical records, whereas 319 patients were recurrence-free according to both methods. The four different criteria all contributed to identifying the patients with recurrence. Most patients fulfilled more than one criterion with the combination of an ICD-10 code for metastasis, a cytostatic therapy code and a SNOMED combination being the most frequent (23 patients). Some patients only fulfilled one criterion (six patients were only identified by a code for metastasis, six by SNOMED combinations only, five by codes for chemotherapy and one by the specific codes for local recurrence).

FIGURE 1

Cohort from previous study.
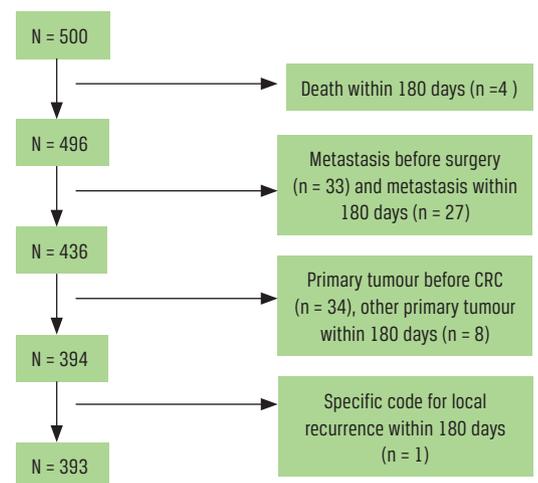
Cohort included in validation analysis

CRC = colorectal cancer.

TABLE 1

Recurrence of rectal cancer identified by the algorithm and by the medical records. The values are n (%).

Algorithm	Medical record		
	recurrence	no recurrence	total
Recurrence	53	14	67 (17.0)
No recurrence	7	319	326 (83.0)
Total	60 (15.3)	333 (84.7)	393 (100)

Kappa statistics showed a high level of concordance between the algorithm and the medical records with a Kappa value (95% confidence interval) of 0.80 (0.72-0.88). The sensitivity and specificity of the algorithm were 88% (77-95%) and 96% (93-98%), respectively.

In seven cases, recurrence was identified through the medical records but was not identified by the algorithm. Furthermore, 14 recurrences were identified by the algorithm but not found in the medical records.

Six patients were identified in the medical records as having recurrent or metastatic disease within the first 180 days but were not excluded by the algorithm. In two of these cases, recurrence was, however, identified after the 180 days by the algorithm.

DISCUSSION

The study showed good concordance between the incidence of recurrence found by the algorithm and that of the medical records review.

The results were similar to those from the study describing the algorithm. In that study, a sensitivity of 95% and a specificity of 97% were found when validating the algorithm [4].

Some discrepancies between the algorithm and the medical records were found. The algorithm identified 14 cases of recurrence that were not identified by the medical records, and the medical records identified seven cases of recurrence that were not identified by the algorithm.

It was possible to find an explanation for some of the 14 recurrences missing from the medical records by reviewing the information from the previous study [8]. Three of these patients were described in the medical records as having metastatic disease before the operation. Thus, the clinicians responsible for the registrations at the time of surgery may have wrongly registered these patients as not having metastatic disease. In one patient, a metastasis to the skin was suspected, but this was never confirmed in the medical record. One could argue that it is a matter of interpretation if this should have been recorded as a recurrence when going through the medical records. Finally, in one patient, a liver metastasis was suspected at first in the medical record, but subsequently the patient was diagnosed with

cholangiocarcinoma. This patient was grouped as having recurrent rectal cancer by the algorithm because the ICD-10 code for metastasis was dated two weeks before the ICD-10 code for cholangiocarcinoma.

The seven cases of recurrence that were identified by the medical records but not by the algorithm may have been missed by the algorithm due to insufficient coding of the clinical findings.

In the present study, medical records were used as the gold standard since this has so far been the method used to identify recurrences after colorectal cancer. All information on the patients from the national electronic medical record was examined. This national electronic medical record should contain the same information as the local medical records at each hospital since information from the local records is automatically uploaded to the national record.

Using the algorithm has several major strengths. Using the algorithm makes it is possible to determine recurrence and disease-free survival for larger groups of patients. So far, it has been possible to determine the incidence of recurrence only by reviewing the patients' medical records. This is time consuming and limits the possible sample size. With the algorithm, it will be possible to perform large register-based studies with recurrence and disease-free survival as primary outcomes, making it possible to assess the effect of changes in various treatment methods. This is valuable, especially in the light of the fact that more patients survive recurrent colorectal cancer for longer periods of time [2, 3]. The algorithm integrates information from various sources, which increases the chances of identifying recurrence even if it has not been identified correctly in one of the registers.

The algorithm also has some design limitations. All patients diagnosed with another cancer before colorectal cancer surgery, except non-melanoma skin cancer, are excluded from the algorithm. This is done because in the registers the diagnosis codes for metastases do not specify which primary cancer caused the metastasis. As survival after other types of cancer also improves, this may prove a challenge in the future. In addition, patients with metastatic disease at the time of surgery or within 180 days after surgery are excluded even if this metastasis has successfully been removed and the patient clinically is considered disease free. It is still possible that a patient was clinically or radiologically diagnosed with a metastasis but that no biopsy was taken, and no treatment was initiated. If the clinician mistakenly did not register an ICD-10 code for metastasis, it would not be identified through the algorithm. Also, the algorithm does not distinguish between local recurrence and distant metastasis. Especially when examining results after rectal cancer surgery, this would have been an advantage.

When using the algorithm, recurrence was defined as metastatic disease or local recurrence after the first 180 days postoperatively. The 180-day limit was chosen arbitrarily. This was done because metastases diagnosed within the first 180 days may represent metastases already present at the time of surgery, but which had not been diagnosed at that time. In the former study, this time limit was not used, and six patients were identified as having recurrent disease within the first 180 days. These patients ought to have been excluded by the algorithm. Two of them were identified by the algorithm as having recurrent disease after the first 180 days. It is to be expected that time of recurrence will be somewhat later when identified by the algorithm than when going through the medical records. Clinical or radiological signs of recurrence are usually mentioned in the medical records before the actual registration of the diagnosis code is done.

The primary strength of this validation study lies in the completeness of access to medical records for the 500 rectal cancer patients from all regions of Denmark. In this manner, it was possible to compare recurrences as they have been determined until now with this new algorithm.

A limitation of the study was that the purpose of reviewing the medical records in the first place was not to validate the algorithm. The decision to validate the algorithm was made after the data extraction for the former study [8] was finished. Although the primary study was performed to identify recurrences, one problem was that when assessing the medical records, it was not explored if the patient had had another primary tumour before his or her colorectal cancer. Ideally, this information would have been available. If this information had been available, it would have been possible to exclude these patients from the medical record cohort separately and then subsequently compare the two groups. Without this information, it was necessary to use only the algorithm to exclude all patients with a primary tumour before colorectal cancer surgery and within the first 180 days.

CONCLUSIONS

The algorithm was found to be suitable and effective in identifying patients with and without recurrence after surgery for rectal cancer. Using this algorithm makes it possible to use recurrence and disease-free survival as endpoints in large register-based studies. Further studies should aim to modify the algorithm so that it may also be used in studies of other types of cancer and to distinguish between local recurrences and distant metastases.

CORRESPONDENCE: *Emilie Palmgren Colov.*
E-mail: eco@regionsjaelland.dk

LITERATURE

1. Hugen N, van de Velde CJH, de Wilt JHW et al. Metastatic pattern in colorectal cancer is strongly influenced by histological subtype. Ann Oncol 2014;25:651-7.
2. Nakajima J, Iida T, Okumura S et al. Recent improvement of survival prognosis after pulmonary metastasectomy and advanced chemotherapy for patients with colorectal cancer. Eur J Cardiothorac Surg 2017;51:869-73.
3. Viganò L, Russolillo N, Ferrero A et al. Evolution of long-term outcome of liver resection for colorectal metastases: analysis of actual 5-year survival rates over two decades. Ann Surg Oncol 2012;19:2035-44.
4. Lash TL, Riis AH, Ostenfeld EB et al. A validated algorithm to ascertain colorectal cancer recurrence using registry resources in Denmark. Int J Cancer 2015;136:2210-15.
5. Pedersen CB, Gøtzsche H, Møller JØ et al. The Danish Civil Registration System. A cohort of eight million persons. Dan Med Bull 2006;53:441-9.
6. Lynge E, Sandegaard JL, Rebolj M. The Danish National Patient Register. Scand J Public Health 2011;39(7 suppl):30-3.
7. Erichsen R, Lash TL, Hamilton-Dutoit SJ et al. Existing data sources for clinical epidemiology: The Danish National Pathology Registry and Data Bank. Clin Epidemiol 2010;2:51-6.
8. Klein M, Colov E, Gögenur I. Similar long-term overall and disease-free survival after conventional and extralevator abdominoperineal excision - a nationwide study. Int J Colorectal Dis 2016;31:1341-7.