Author/s:
MacCallum, C;Da Silva, N;Gibbs, P;Thomson, BNJ;Skandarajah, A;Hayes, I

Title:
Accuracy of administrative coding data in colorectal cancer resections and short-term outcomes

Date:
2018-09-01

Citation:
MacCallum, C., Da Silva, N., Gibbs, P., Thomson, B. N. J., Skandarajah, A. & Hayes, I. (2018). Accuracy of administrative coding data in colorectal cancer resections and short-term outcomes. ANZ Journal of Surgery, 88 (9), pp.876-881. https://doi.org/10.1111/ans.14714.

Persistent Link:
https://hdl.handle.net/11343/284363

Da Silva Nigel (Orcid ID: 0000-0003-3522-7419)

1. Title Page

A. Title

ACCURACY OF ADMINISTRATIVE CODING DATA IN COLORECTAL CANCER RESECTIONS AND SHORT TERM OUTCOMES.

B. Authors and qualifications

Caroline McCallum[1,2,4*]; **Nigel Da Silva**[1,2,5*]; Peter Gibbs[3,6]; Benjamin NJ Thomson[2,7]; Anita Skandarajah[2,8]; Ian Hayes[1,2,9]

[4] MBBS, MS
[5] FRACS
[6] FRACP, MD
[7] FRACS, DMedSc
[8] FRACS, MD
[9] FRACS, MS

* Both authors contributed equally to this paper

C. Authors Affiliation(s)

[1] Colorectal Surgery unit, Royal Melbourne Hospital (RMH),
[2] Department of General surgical specialities, RMH
Department of surgery, The University of Melbourne
[3] Walter and Eliza Hall Institute of Medical Research, Melbourne; The University of Melbourne

D. Corresponding author

Name: Nigel Da Silva
Address: 18/35 Victoria Parade, Mawson lakes, SA 5095
Email: nigel.dasilva@gmail.com
Mob: 0413724626

Not recipient of a scholarship

2. Abstract

Background

Administrative data are routinely captured for each hospital admission and may serve as an alternative source for populating databases. This study aims to determine the accuracy of administrative data to provide tumour characteristics and short term post-operative outcomes, after a colorectal cancer resection, compared with clinical data.

Methods

A retrospective study of all colorectal cancer resections at a single hospital from 1st of January 2008 to 31st of December 2013 was conducted. Local administrative data was coded as per ICD-10 AM (International Classification of Diseases, Tenth Revision, Australian Modification) and ACHI (Australian Classification of Health Interventions). Clinical data for all patients was extracted from the medical charts and compared with administrative data. Code combinations and algorithms were used to improve the accuracy of administrative data.

Results

A total of 436 patients were identified. The accuracy of algorithms combining tumour location and type of operation for right colon, left colon and rectum were 93%, 89% and 88% respectively. The accuracy of histological type was 89%, lymph node status 92% and metastasis status 88%. The accuracy of return to theatre and in-hospital mortality was 100%.

Conclusion

Administrative data can provide reliable information on tumour details and short-term post-operative outcomes. The potential for administrative data to validate data captured in registries and be used independently for audit and research should be further explored.

3. Main Text

Introduction

Colorectal cancer (CRC) is the second most commonly diagnosed cancer in Australia. The estimated number of new cases of colorectal cancer diagnosed in 2017 is 16,682, comprising 12.4% of all new cancer cases [1]. Clinical colorectal cancer registries (CCCR) are potentially powerful tools in the study of cancer and can positively impact on the value, efficiency and resource allocation in the management of CRC. However, they are resource intensive, region specific, and compromised by reporting bias [2].

Hospital administrative data, which are routinely collected, are an alternative source of tumour, treatment and outcome data [3, 4]. They are readily available, cheap and have a broad geographical scope[5].

However, the accuracy of administrative data has not been fully explored and may vary by data item. Outcomes are limited to the available data fields, which may be less extensive than the information captured in a clinical registry [4]. The quality of the administrative coding system is highly standardised. A recent study on administrative data in acute diverticulitis has demonstrated good correlation with clinical notes for a non-malignant condition [6]. Here we explore data for a common cancer.

The aim of this study was to determine the accuracy of administrative data to provide tumour characteristics and short term post-operative outcomes, after a colorectal cancer resection, compared with clinical data.

Methods

A retrospective analysis of all colorectal cancer (CRC) resections from 1st of January 2008 to 31st of December 2013 was conducted at the Royal Melbourne Hospital (RMH), Victoria, Australia.

Initially a cohort of RMH patients with CRC, excluding duplicates and prevalent cases, was obtained from the RMH component of the BioGrid Australia CRC database (Figure 1). BioGrid Australia was established as a technology platform to enable linkage and analysis of de-identified data from disparate databases without compromising data security or patient confidentiality, in an ethically approved manner [7]. A similarly defined list of patient episodes was generated from the RMH administrative database. Our first step was to match these two lists.

We anticipated some discrepancy between the cohorts identified using the two databases [8]. There was a broad pick-up of cases from the administrative database including cases subsequently transferred to private health care facilities for operative management and emergency admissions for palliative care or symptomatic management. In order to obtain a better match between the two databases it was necessary to redefine and restrict the cohort to colorectal cancer resection cases. Cohort-defining algorithms were constructed using combinations of principal diagnosis, resection and type of admission codes. The best matched algorithm (Figure 1) was subsequently applied to the RMH BioGrid CRC database and the RMH administrative database. Reasons for non matched cases were elicited from the medical charts of each patient. Clinical data was extracted and tabulated from each patient's complete medical case notes and the hospital clinical information system (CIS). Chart review revealed that clinical documentation was not always consistent, particularly in relation to the anatomical site specified in the discharge summary and operation note. If there was a discrepancy, the operation note was used as the prime source for identifying both the anatomical site and operation performed. Administrative data was derived from the respective administrative coding sheets.

The nine data fields tested for accuracy were anatomical tumour location, type of resection, histology, lymph node status, metastasis status, in-patient mortality, return to theatre, post-operative transfusion requirement and post-operative wound infection. All of these data fields are coded as per the ICD-10-AM and ACHI codes, in the administrative database (table 1).

Combinations of codes were used to improve the accuracy of anatomical tumour location. Anatomic locations were grouped into right colon, left colon and rectum. Right colon was defined as extending to and including the splenic flexure. Left colon was comprised of the descending colon and sigmoid colon. Rectum included rectosigmoid and rectum. These division were done based on anatomical and oncologic differences [9, 10]. Further algorithms were then constructed to combine type of operation and tumour location in three divisions: right colon, left colon and rectum (Table S1). Tumour location was the primary determinant of grouping. Thus, extended right hemicolectomy for a transverse colon tumour was grouped in the right colon group; high anterior resection for a sigmoid cancer was grouped in the left colon group. In a small number of cases, coding combinations were illogical (e.g., abdominoperineal resection coded for a caecal cancer). These cases could not be grouped into any of the three divisions and were thus considered inaccurate.

ICD-10-AM only codes the site of lymph node involvement but does not contain information regarding the number of lymph nodes involved with tumour. Coding for metastasis status includes the site of metastasis. To simplify data comparison, metastasis status was recorded as the presence or absence of non-lymph node metastasis. The accuracy of liver metastasis was also studied.

'Return to theatre' was used for an unplanned operation following the initial colorectal surgery, during the same admission. The algorithm used to extract this information is noted in Table 1.D.

'In-hospital mortality' referred to a patient dying in hospital following the initial colorectal surgery, during the same admission. This was recorded in the 'destination' data field, which recorded the destination of the patient following their hospital admission. The inclusion criteria for 'post-operative anaemia requiring transfusion', is noted in Table 1.E. Episodes of post-operative wound infection were coded as per ICD-10-AM (Table 1.F).

Statistical test

Data fields were tested for accuracy. Accuracy denoted the closeness of the measured value to the actual value. It was calculated as a percentage: accuracy= [(true positives+ true negatives)/ (true positives + true negatives + false positives+ false negatives)] * 100.

Results

An initial broad search strategy for all cases of colorectal cancer in RMH administrative data and Biogrid produced a cohort of 694 patients asymmetrically distributed between the two databases. A more refined search restricting the cohort to resected cases produced a more balanced and more clinically relevant case-mix. A total of 465 patients were eligible for final analysis. The cohort-defining algorithm identified 433 matches between the databases, with 23 unique cases in the administrative database not identified in the BioGrid database and 16 unique cases in the BioGrid database not identified in the administrative database. Of the unique cases, 32 fulfilled the inclusion criteria. Medical charts were unavailable for 29 cases and hence the final cohort included 436 patients.

Individual and code-combination anatomical site accuracies are noted in Table 2. The position of cancers at the hepatic flexure and the rectosigmoid was poorly coded, with an accuracy of 53% and 64% respectively. By grouping anatomical locations into right sided, left sided and rectal combinations, accuracy of position was further improved to 94% and 90% and 88% respectively.

The overall accuracy for operation coding was 71%. Individual operation coding accuracies are noted in Table 3. Rectosigmoidectomy with formation of stoma was consistently coded with a high accuracy of 94%. High anterior resection of rectum had a lower accuracy of 59%.

The accuracy of the final combinations of right colon, left colon and rectal operation and associated tumour location, with the constructed algorithms were 93%, 89% and 88% respectively (Figure S1).

The overall accuracy of histology was 89%. The dominant histological type was adenocarcinoma, which was coded with 95% accuracy. In 23 cases, predominantly rectal cancers following neoadjuvant treatment, there was 'no residual tumour' in the resected specimen, so adenocarcinoma was not necessarily recorded in the pathology report. Such cases were recorded as adenocarcinoma in 14 of the 23 cases. The accuracy of lymph node and distant metastasis status was 92% and 88% respectively. Liver metastasis as an isolated entity was coded with an accuracy of 42%.

Return to theatre and in-patient mortality had an accuracy of 100%, i.e., 11 and 10 cases respectively. The accuracy of post-operative wound infection was 67% and post-operative anaemia requiring blood transfusion was 81%.

Discussion

Administrative data are increasingly being used to explore surgical outcomes in research and audit. Clinical interpretation of administrative data requires a clear cohort definition [4]. Combinations of data fields and algorithms can be used to define more specific cohorts [11].

In this study, the accuracy of administrative data for tumour characteristics and short-term post-operative outcomes, after a colorectal cancer resection, were compared with clinical data. The nine data fields were chosen because they are important determinants of surgical treatment and post-operative outcomes and because they were extractable from administrative data.

The process of identifying the complete cohort of colorectal cancer patients who had been treated at our hospital proved to be more complicated than anticipated. Our intention had been to use Biogrid as a starting point for case finding. Biogrid is an excellent database of clinical material but its case accrual methods, through scrutiny of colorectal surgery unit outpatient clinics and unit theatre lists, can miss emergency cases and cases treated by other units. Field et al. in a 2010 study, compared the accuracy of colorectal cancer patient identification between the Victorian Cancer Registry (VCR) and the BioGrid database [8]. The results were similar to this study, with the Biogrid database missing 91 of 913 cases identified by VCR between 2000 and 2005.

In contrast, administrative data only captures inpatient admissions. Administrative data can be over-inclusive in picking up patients who have colorectal cancer in their list of diagnoses but are admitted for other reasons. Such patients may not come to the attention of the colorectal surgical service. Administrative data picked up more cases than Biogrid.

There was significant variation in the accuracy of coding of anatomical site and type of resection. The overall accuracy of tumour location was 59% & the overall accuracy for operation coding was 71%. We hypothesized that algorithms of codes could be used to increase accuracy of colorectal cancer coding. This method was used in a similar study on diverticulitis [11]. Algorithms were derived from observations regarding the coding errors, and from logical combinations of codes according to anatomical and clinical relevance. Errors in administrative or clinical databases can have three aetiologies: errors in the medical records that were translated into the database, errors of interpretation of medical records, and errors of data entry [12]. An important observation from our study was that the inaccuracies of anatomical position and operation type often involved only small alterations in location (e.g. caecum/ascending colon). Thus, it was possible to combine anatomical sites sharing similar lymphatic drainage and operation type, with the operation codes, into groups that were logical from a surgical oncology perspective. The accuracy of the final algorithms combining operation and tumour location codes for right colon, left colon and rectal groups were 93%, 89% and 88% respectively. These grouping also rank the sites and surgical procedures from least to most technically difficult. Although these groupings lose some of the detail of the data, they allow the initially unpromising accuracy figures about tumour location to regain clinical utility. The accuracy level achieved with this is commensurate with the data accuracy level found in several international cancer registries including New Zealand's breast cancer registry, the Geneva cancer registry and the national program of cancer registries assessment of SEER summary stage in US colorectal cancer registries [13-15].

The accuracy of histology, lymph node status and metastasis status were 89%, 92% and 88% respectively. However, liver metastasis as an isolated entity was coded with an accuracy of only 42%. Post-operative outcomes e.g., return to theatre and in-patient mortality had an accuracy of 100%.

For histological type, lymph node status and metastasis status there was good coding accuracy without the requirement for algorithms. Staging is not formally described in administrative data but enough information was found in this study to allow tumour staging to be approximated.

In-hospital mortality and return to theatre were coded with 100% accuracy without the need for coding algorithms. Return to theatre only included subsequent operations that related directly to the index procedure; for instance, it did not include an orthopaedic procedure for a fractured limb secondary to a post-operative fall during the admission.

It should be noted that there is no administrative code for an anastomotic leak, but classification of anastomotic leak, even from clinical notes, can be problematic. In a 2013 study by Reinke et al., the authors evaluated the ability of 2 sources, an administrative database and a clinical registry to identify anastomotic leaks [16]. Due to a lack of a definition for anastomotic leak both the administrative database and clinical registry, logical proxies were used. Although the clinical registry had higher sensitivity and specificity for anastomotic leak, both databases had low sensitivity.

In coding for post-operative anaemia, coding could not differentiate between pre-existing anaemia and anaemia caused by post-operative haemorrhage. Both these types of anaemia are associated with poorer surgical outcomes but this is a limitation [17]. The poor recording of post-operative wound infection is not surprising, as the diagnosis is somewhat subjective and unlike other events measured, may not be well documented in clinical notes.

A limitation of this study is that only one tertiary-level hospital, RMH, was used for the analysis. It is possible that coding practices could vary across hospitals, affecting the generalisability of these results. However, Australian coders must adhere to the Australian Coding Standards; this adds predictability to their coding of admissions and suggests that coding practices are likely to be similar among hospitals. Our group has previously used a methodology similar to this study to examine administrative coding for acute diverticulitis by comparing it with RMH case note clinical data. Algorithms of codes for diverticulitis were created and applied to 8 tertiary referral hospitals in Victoria. The diverticular coding showed similar patterns of disease and treatment across all hospitals, supporting the concept that coding behaviour was consistent across these institutions. [11]. Furthermore, the sample size of this current analysis was large enough to account for common coding errors and practices.

Conclusion

This study quantifies the accuracy of administrative data for colorectal cancer resection by comparison against clinical notes. We found that administrative data provide reasonably detailed information concerning tumour histology, staging, location, surgical procedure and post-operative complications. The coding of administrative data was well correlated with clinical notes for return to theatre, in-patient mortality, lymph node status, metastasis status and histology. Tumour location and procedure were coded with variable accuracy which was improved to approximately 90% by grouping data using algorithms. Administrative data is not identical to clinical data but this study reveals a reasonable level of accuracy and gives some indication of the potential to utilise this huge resource in the study of colorectal cancer.

4. Acknowledgements

The author is grateful to Leonie Dodds, acting coding manager, Royal Melbourne Hospital for providing the ICD-10-AM and ACHI codes

5. Disclosure statement

The author declares no conflict of interest

6. References

1. 2017., A.I.o.H.a.W.A. *Australian Cancer Incidence and Mortality (ACIM) books: Colorectal cancer*. 2017 03 Feb 2017; 2017:[Available from: http://www.aihw.gov.au/acim-books.
2. Graves, S.E., *The value of arthroplasty registry data.* Acta Orthop, 2010. **81**(1): p. 8-9.
3. Bennett, I.C., *Surgeons and big data.* ANZ J Surg, 2017. **87**(9): p. 647-648.
4. Hong, M.K., A.R. Skandarajah, and I.P. Hayes, *Administrative data: what surgeons should know about big data.* ANZ J Surg, 2017. **87**(9): p. 650-651.
5. Parkin, D.M., *The evolution of the population-based cancer registry.* Nat Rev Cancer, 2006. **6**(8): p. 603-12.
6. HONG, M.K.-Y.e.a., *How to use administrative data for surgical outcomes in acute diverticulitis.* General Surgery, 2016. **V.1**: p. 5-12.
7. Kosmider, S., et al., *Towards establishing a national colorectal cancer database: lessons learnt from Bio21 molecular medicine informatics model.* ANZ J Surg, 2008. **78**(9): p. 803-9.

8. Field, K., et al., *Linking data from hospital and cancer registry databases: should this be standard practice?* Intern Med J, 2010. **40**(8): p. 566-73.

9. Suttie, S.A., et al., *Outcome of right- and left-sided colonic and rectal cancer following surgical resection.* Colorectal Dis, 2011. **13**(8): p. 884-9.

10. RA, M., et al., *Is there a difference in survival between right- versus left-sided colon cancers?* Ann surg oncol 2008. **15**(9): p. 2388-94.

11. Hong, M.K., et al., *Operative intervention rates for acute diverticulitis: a multicentre state-wide study.* ANZ J Surg, 2015. **85**(10): p. 734-8.

12. Goldberg, S.I., A. Niemierko, and A. Turchin, *Analysis of data errors in clinical research databases.* AMIA Annu Symp Proc, 2008: p. 242-6.

13. Thoburn, K.K., et al., *Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries.* Cancer, 2007. **109**(8): p. 1607-16.

14. Schaffar, R., et al., *Accuracy of cause of death data routinely recorded in a population-based cancer registry: impact on cause-specific survival and validation using the Geneva Cancer Registry.* BMC Cancer, 2013. **13**: p. 609.

15. Seneviratne, S., et al., *Accuracy and completeness of the New Zealand Cancer Registry for staging of invasive breast cancer.* Cancer Epidemiol, 2014. **38**(5): p. 638-44.

16. Reinke, C.E., et al., *Comparison of anastomotic leak rate after colorectal surgery using different databases.* Dis Colon Rectum, 2013. **56**(5): p. 638-44.

17. Smilowitz, N.R., et al., *Association Between Anemia, Bleeding, and Transfusion with Long-term Mortality Following Noncardiac Surgery.* Am J Med, 2016. **129**(3): p. 315-23 e2.

7. Figure and table Legends

1. Figure 1. Cohort search
2. Table 1. ICD-10-AM & ACHI codes (independent Hospital Pricing Authority ICD-10-AM/ACHI 7th & 8th edition)
3. Table 2. Accuracy of administrative data for anatomical location of tumour
4. Table 3. Accuracy of administrative data for type of operation before algorithms

8. List of supporting information

1. Table S1. Algorithms of combined operation type and anatomical tumour position
2. Figure S1. Accuracy of algorithms of combined operation type and anatomical position

9. Tables

1. Table 1. ICD-10-AM & ACHI codes (independent Hospital Pricing Authority ICD-10-AM/ACHI 7th & 8th edition)
2. Table 2. Accuracy of administrative data for anatomical location of tumour

3. Table 3. Accuracy of administrative data for type of operation before algorithms

   (Attached as separate file)

   (750 words)

10. Figures

1. Figure 1. Cohort search

(Attached as separate file)

(250 words)

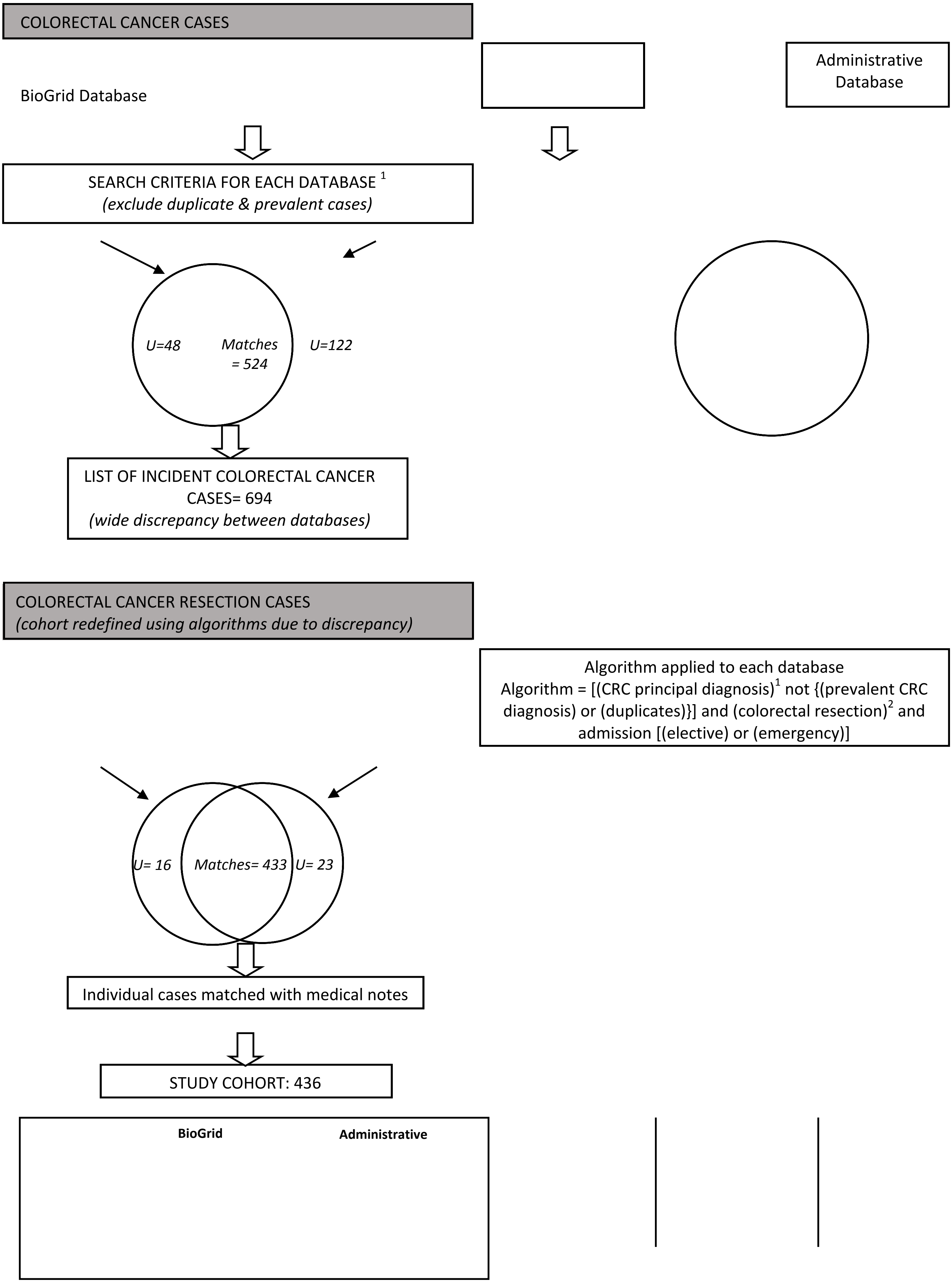
11. Total word count

a. Abstract: 223 words

b. tables and figures: 1000 words

c. total words: 3990

Figure 1. Cohort search



[1] **Principal diagnosis**	'diagnosis data' field for CRC	ICD-10-AM codes C180-C189, C19, C20
[2] **Colorectal resection**	'operation' data field for CRC resection	ACHI codes

U= unique cases; ICD-10-AM & ACHI codes from table 2 & 3 respectively

Table 1. ICD-10-AM & ACHI codes *(Independent Hospital Pricing Authority ICD-10-AM/ACHI 7th & 8th edition)*

A. Histopathology	
ICD-10-AM	**Description**
M81403 & M84803	Adenocarcinoma
M82633	Adenocarcinoma in tubulovillous adenoma
B. Lymph node status	
C772	Secondary and unspecified malignant neoplasm of intra-abdominal lymph nodes
C775	Secondary and unspecified malignant neoplasm of intrapelvic lymph nodes
C. Metastasis status (description denotes metastasis to that organ)	
C780	Secondary malignant neoplasm of lung
C782	Secondary malignant neoplasm of pleura
C786	Secondary malignant neoplasm of retroperitoneum and peritoneum
C787	Secondary malignant neoplasm of liver and intrahepatic bile duct
C796	Secondary malignant neoplasm of ovary
C7988	Secondary malignant neoplasm of other specified sites
D. Return to theatre Algorithm= *(General anaesthetic codes as above in section A) + [(Postoperative reopening of laparotomy site-3038500) or (other operation directly relating to index colorectal resection: ACHI codes noted in study are below)]*	
3201200	Total colectomy and ileorectal anastomosis
3037300	Exploratory laparotomy
3037529	Temporary ileostomy
3200000*	Limited exciision of large intestine with formation of stoma
E. Post-operative anaemia requiring transfusion (ICD-10-AM) Algorithm= *procedure code: administration of packed cells (1370602) + one diagnosis code (as below)*	
D64.9	Anaemia, unspecified
D62	Acute posthaemorrhagic anaemia
T81.0	Haemorrhage and haematoma complicating a procedure, not elsewhere classified
K92.2	Gastrointestinal haemorrhage, unspecified
D50.9	Iron deficiency anaemia, unspecified
D63.8	Anaemia in other chronic diseases classified elsewhere
D63	Anaemia in neoplastic disease
F. Post-operative wound infection	
T81.4	Wound infection following a procedure, not elsewhere classified
T81.41	Wound infection following a procedure

Table 2. Accuracy of administrative data for anatomical location of tumour

Anatomical site	ICD-10-AM codes	Accuracy	Code combinations (without algorithms)
Appendix	**C181**	**100%**	**RIGHT COLON**
Caecum	**C180**	**84%**	
Ascending colon	**C182**	**79%**	
Hepatic flexure	**C183**	**53%**	} **94%**
Transverse colon	**C184**	**87%**	
Splenic flexure	**C185**	**80%**	
Synchronous caecal & splenic flexure	**C218**	**100%**	
Descending colon	**C186**	**71%**	**LEFT COLON**
Sigmoid	**C187**	**85%**	} **90%**
Rectosigmoid	**C19**	**64%**	**RECTUM**
Rectum	**C20**	**79%**	} **88%**

Table 3. Accuracy of administrative data for type of operation before algorithms

Operation	ACHI codes	Accuracy
Appendicectomy	3057001	100%
Right hemicolectomy with anastomosis	3200301	85%
Extended right hemicolectomy with anastomosis	3200501	67%
Extended right hemicolectomy with formation of stoma	3200401	100%
Subtotal colectomy with anastomosis	3200500	86%
Subtotal colectomy with formation of stoma	3200400	60%
Left hemicolectomy with anastomosis	3200600	86%
Left hemicolectomy with formation of stoma	3200601	67%
Rectosigmoidectomy with formation of stoma	3203000	94%
High anterior resection of rectum (HAR)	3202400	59%
Total colectomy with ileorectal anastomosis	3201200	83%
Total colectomy with ileostomy	3200900	100%
Total proctocolectomy ileoanal anastomosis and stoma	3205101	100%
Total proctocolectomy with ileostomy	3201500	100%
Low anterior resection of rectum	3202500	66%
Ultra low anterior resection of rectum	3202600	79%
Abdominoperineal proctectomy	3203900	83%
Limited excision of large intestine with anastomosis	3200300	19%
Ileocolic resection	3051503	0%
Resection of small intestine with anastomosis	3056600	100%