Can Text-Search Methods of Pathology Reports Accurately Identify Patients with Rectal Cancer in Large Administrative Databases?

Reilly P. Musselman[1], Deanna Rothwell[2], Rebecca C. Auer[1], Husein Moloo[1], Robin P. Boushey[1], Carl van Walraven[2]

[1]Division of General Surgery, University of Ottawa, [2]Department Epidemiology and Community Medicine, Ottawa Hospital Research Institute, Ottawa, ON, Canada

Abstract

Background: The aim of this study is to derive and to validate a cohort of rectal cancer surgical patients within administrative datasets using text-search analysis of pathology reports. **Materials and Methods:** A text-search algorithm was developed and validated on pathology reports from 694 known rectal cancers, 1000 known colon cancers, and 1000 noncolorectal specimens. The algorithm was applied to all pathology reports available within the Ottawa Hospital Data Warehouse from 1996 to 2010. Identified pathology reports were validated as rectal cancer specimens through manual chart review. Sensitivity, specificity, and positive predictive value (PPV) of the text-search methodology were calculated. **Results:** In the derivation cohort of pathology reports ($n = 2694$), the text-search algorithm had a sensitivity and specificity of 100% and 98.6%, respectively. When this algorithm was applied to all pathology reports from 1996 to 2010 ($n = 284,032$), 5588 pathology reports were identified as consistent with rectal cancer. Medical record review determined that 4550 patients did not have rectal cancer, leaving a final cohort of 1038 rectal cancer patients. Sensitivity and specificity of the text-search algorithm were 100% and 98.4%, respectively. PPV of the algorithm was 18.6%. **Conclusions:** Text-search methodology is a feasible way to identify all rectal cancer surgery patients through administrative datasets with high sensitivity and specificity. However, in the presence of a low pretest probability, text-search methods must be combined with a validation method, such as manual chart review, to be a viable approach.

Keywords: Administrative databases, pathology reports, rectal cancer, text search

Introduction

With progressive digitization of healthcare and enhanced computing power, administrative data have become an increasingly available resource for clinical researchers. The use of administrative databases for healthcare-related research publications has risen steadily over the past two decades, with results that often lead to important conclusions.[1] In administrative data, disease cohorts are primarily identified using diagnostic or procedural codes.[2]

Several problems exist regarding the use of healthcare-related codes to identify patient cohorts within administrative datasets. A wide range of the accuracy of administrative data codes has been reported.[3,4] There frequently appears to be no accepted standard collection of codes to identify particular diseases, for example, Hohl *et al.* (2014) demonstrated extensive variability between studies in the numbers and types of codes used to identify the same population.[5] In addition, very few studies actually validate the codes they use to create datasets for analysis.[2] This indicates that the accuracy of the surrogate (the code) for the entity it is supposed to represent is unknown. As such, the generalization of results using the code to those involving the entity that is trying to be studied is unreliable.[6]

The use of administrative database diagnostic and procedural codes for cohort identification can be particularly problematic

Address for correspondence: Dr. Reilly P. Musselman,
The Ottawa Hospital, Civic Campus, 737 Parkdale Avenue, Room CPC 300,
Ottawa, ON K1Y 1J8, Canada.
E-mail: rmusselman@toh.ca

Access this article online

Quick Response Code:

Website:
www.jpathinformatics.org

DOI:
10.4103/jpi.jpi_71_17

when studying patients with rectal cancer. Despite clear differences in both outcomes and treatment protocols, differentiating rectal from colon cancer patients using administrative codes can be extremely difficult.[7-12] One reason for this is the varying definitions that can be used for rectal cancer including tumors that are within 15 cm of the anal verge;[13] lie below the peritoneal reflection of the abdominal cavity (i.e., the division between the abdominal cavity and the pelvis);[14] are amenable to radiation therapy;[15] or are described as "rectal cancers" by the clinician.[16] The range of potential rectal cancer definitions can make it difficult to create a homogeneous population of patients, especially in retrospective studies where the inclusion criteria cannot be made *a priori*. The patient misclassification resulting from these varying clinical definitions would be amplified in studies using health administrative data in which cases are identified using diagnostic codes. In this situation, error associated with the clinical classification of rectal cancer is amplified by error associated with inaccurate diagnostic codes that are assigned based on medical record review by a health records analyst.

Because of these issues, novel ways to create reliable patient cohorts are necessary to study rectal cancer using health administrative data. Text-analytical methods using radiology and pathology reports have been successfully utilized in the study of cardiac and breast cancer patients.[17,18] To date, no published data on the use of text-search methods to identify rectal cancer surgery cohorts are available. The purpose of the present study was to derive and to validate a cohort of rectal cancer surgery patients using a text-search algorithm of pathology reports from 1996 to 2010 at a single, tertiary-care institution.

MATERIALS AND METHODS

Definition of rectal cancer

This study defined rectal cancer as "an adenocarcinoma of the rectum for which the distal margin of the tumor was within 15 cm of the anal verge." This definition is consistent with that used in major clinical trials for the disease.[12,13,19,20] Distance measurements based on rigid sigmoidoscopy were preferentially used. If such measurements were unavailable from medical records, measurements from the anal verge were based on the following sources (in order of preference): Flexible endoscopy; magnetic resonance imaging; and digital rectal examination. If no distance measurement could be found from any of these sources, rectal cancer status was determined using best clinical judgment from information available in the medical record.

Datasets

The Ottawa Hospital Data Warehouse (OHDW) is a relational database containing data from all operational information systems at The Ottawa Hospital (TOH), a tertiary care institution. Operational information systems include clinical data such as laboratory results, clinical notes, pathology reports, hospitalization discharge abstract records, and demographic data.

This study also used the TOH-Colon and Rectal Cancer (CRC) registry. This database included all patients who were coded with Canadian Classification of Health Intervention procedure codes indicating colorectal surgery between 2002 and 2010 at TOH. The accuracy and completeness of this registry is unknown.

Development of text-search algorithm

With permission from our research ethics board, we extracted from the TOH-CRC all patients and surgery dates of rectal cancer resections [Figure 1-A]. The medical records for each of these encounters were reviewed by RM. Patients meeting inclusion criteria listed in Table 1 were classified with rectal cancer and were included in the study. This dataset of rectal cancer patients was uploaded to the OHDW, wherein the health record number was encrypted to permit anonymous linkage with patient-specific records within the OHDW [Figure 1-B].

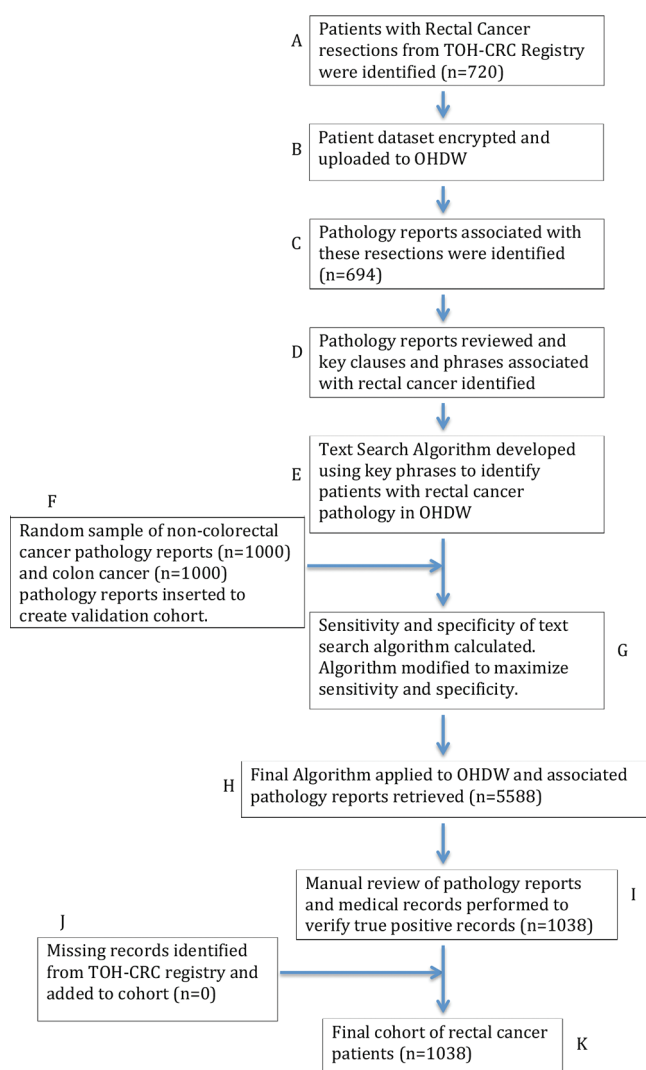**Figure 1:** Flow diagram depicting methodology used to identify all rectal cancers at our institution. Letters A-G represent specific steps used to create the cohort and are referenced in the study's text. Letters H-K respresent steps taken to validate the text-search algorithm of pathology reports through manual chart review. TOH-CRC: The Ottawa Hospital-Colorectal Cancer, OHDW: Ottawa Hospital Data Warehouse

Table 1: Inclusion and exclusion criteria for surgically treated rectal cancer patients

Inclusion criteria
    Diagnosis of rectal cancer
    (adenocarcinoma with distal tumor margin ≤15 cm of anal verge)
    Age ≥18 years
    Resection of rectal cancer with nonpalliative surgery between 1996 and 2010
    Pathology reports, hospitalization records, and outpatient clinic notes available
Exclusion criteria
    Cancer of the colon (based on pathology reports and clinical notes)
    Rectal cancer with palliative resection

We linked the colorectal cancer patient database to the repository of pathology reports using patient unique identifiers and date of surgery to retrieve pathology reports for all rectal cancer cases in TOH-CRC [Figure 1-C]. These pathology reports were manually reviewed by RM to identify potential clauses and phrases that were repeated frequently that might be used in a text-search algorithm to identify pathology reports of rectal cancers [Figure 1-D]. Once identified, different iterations of these words, clauses, and phrases within the text-search macro were tested for their ability to accurately identify rectal cancer resections.

We applied these phrases to a SAS macro (SAS Institute, Cary, NC, USA) that analyzes text for clauses of interest with or without preceding or following specified qualifiers [Figure 1-E]. This SAS macro facilitates computerized text analysis of clinical reports to identify different disease processes or diagnoses within large data warehouses and has been previously described and validated for this use.[21] Search terms identified in step 1D were then applied in several iterations (i.e., using slightly different versions of the words or terms and in different order each time) to the dataset from Figure 1-C to create a text-search algorithm using terms as specific as possible while maintaining a sensitivity of at least 99% to ensure we identified essentially all surgically treated rectal cancer patients at our institution. This was termed the preliminary rectal cancer pathology text-search algorithm.

To modify the preliminary algorithm, we identified a large sample of pathology reports for patients without rectal cancer [Figure 1-F]. First, 1000 random pathology reports were extracted from the OHDW. These included surgical specimens and biopsy results from a variety of procedures. These reports were reviewed to ensure no colon or rectal pathology was included. Then, 1000 colon cancer reports were identified using the TOH-CRC and added to this population. Finally, we added the pathology reports from the 694 rectal cancer patients identified in Figure 1-E. This resulted in a total of 2694 pathology reports on which we tested our text-search algorithm. This population of pathology reports was termed the "validation cohort."

We applied the text-search algorithm to the validation cohort of pathology reports containing diagnoses of rectal cancer, colon cancer, and noncolorectal cancer specimens. The algorithm was then further modified to maximize sensitivity and specificity until a final rectal cancer pathology text-search algorithm was identified [Figure 1-G].

Text-search of all pathology reports

The final rectal cancer pathology text-search algorithm was then applied to all pathology reports stored in the OHDW [Figure 1-H]. All pathology reports identified by the rectal cancer pathology text-search algorithm were then manually reviewed with their associated operative notes, clinic notes, radiology reports and hospitalization discharge summaries to determine if they met inclusion criteria for rectal cancer listed in Table 1 and Figure 1-I. Those not meeting inclusion criteria outlined in Table 1 were considered to be false positive rectal cancer pathology text-search algorithm results and were excluded from the rectal cancer cohort. Finally, we linked this cohort of rectal cancer patients to the original TOH-CRC registry to determine if any rectal cancer cases were missed [Figure 1-J]. This yielded the final cohort of patients who underwent a surgical resection for rectal cancer at TOH between 1996 and 2010 [Figure 1-K]. Sensitivity, specificity, and positive predictive value (PPV) for the text-search algorithm were calculated.

RESULTS

Seven hundred and twenty patients with a surgical resection for rectal cancer were identified in the TOH-CRC registry [Figure 1-A]. After manual review of patient records, 694 met study inclusion criteria [Figure 1-C].

These pathology reports were reviewed to identify keywords and phrases potentially indicating rectal cancer. Multiple iterations of the text-search algorithm were then attempted to maximize the identification of known rectal cancer cases. A total of 7 iterations of the text-search algorithm were applied to the pathology reports of the 694 known rectal cancer cases. Each text-search iteration utilized different keywords and phrases, with the best sensitivity score being 100%. The third iteration was selected as the final rectal cancer pathology report text-search algorithm because it maintained 100% sensitivity while using search terms that were qualitatively most specific to rectal cancer resections. This iteration was used as the preliminary rectal cancer pathology text-search algorithm.

This preliminary algorithm was applied to a validation cohort of pathology reports that contained the 694 known rectal cancers, 1000 known colon cancer resections, and 1000 randomly selected noncolon, nonrectal cancer patients. A total of five iterations of the text-search algorithm modifications were created [Table 2]. Different iterations were tested to find the combination of words and phrases that maintained 100% sensitivity while maximizing specificity, producing the final rectal cancer pathology text-search algorithm. This final

Table 2: Second round results of text-search algorithm* development

Iteration number	Keyphrases leading to study inclusion**	Keyphrases leading to study exclusion¶	Sensitivity (%)	Specificity (%)
1	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	None	100	74.2
2	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Colo" or "Colon")	62.2	99.6
3^	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Colo" or "Colon") NOT within 20 spaces of ("Rectum" or "Rectal")	98.6	95.0
4	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	[("Colo" or "Colon") NOT within 20 spaces of ("Rectum" or "Rectal")] AND within 20 spaces of ("Diagnosis")	100	91.2
5^	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	[("Colo" or "Colon") NOT within 20 spaces of ("Rectum" or "Rectal")] AND within 20 spaces of ("Specimen" or "Diagnosis")	100	98.6

*Algorithm development is based on a modified version of a previously described software code template called a "macro" using SAS statistical software, version 9.2 (NC, USA), **These represent key words and phrases included in the algorithm that, when identified within the pathology report, would lead to a screen-positive, and therefore, inclusion of that pathology report into the study, provided no exclusion phrases are found, ¶These represent key words and phrases that, if identified within the pathology report, lead to automatic exclusion from the study, regardless of whether or not key inclusion words/phrases are identified, ^Combination of key words/phrases used for the final text-search algorithm. Each iteration was performed on pathology reports from a population of 694 known rectal cancer resections performed at The Ottawa Hospital, along with 1000 known colon cancer resections, and 1000 random, noncolorectal cancer cases. Sensitivity: Calculated by dividing the number of true positives captured in the test by the total number of true positives (*n*=694), Specificity: Calculated by dividing the number of true negatives captured in the test by the total number of true negatives (*n*=2000)

algorithm produced a screening test that was 100% sensitive and 98.6% specific [Table 2].

All pathology reports available at TOH between January 1, 1996 and December 31, 2010 were extracted from the OHDW (*n* = 284,032) [Figure 2-A]. The final rectal cancer pathology text-search algorithm was applied and a total of 5588 pathology reports were identified by final text-search algorithm [Figure 2-B]. The medical records associated with these pathology reports were reviewed to determine those meeting final criteria for rectal cancer [Table 1]. Of the 5588 screen-positive reports between January 1, 1996 and December 31, 2010, 4106 were excluded on the first-pass reading of the pathology reports themselves [Figure 2-C]. This left 1482 for manual chart review of medical records, which included operative reports, clinic notes, and radiology reports. This process excluded another 444 patients, leaving a total of 1038 patients who met inclusion criteria of the present study [Figure 2-D]. All 694 cases initially identified from the TOH-CRC registry were included in this final cohort of 1038 rectal cancers.

Assuming a sensitivity of 100%, the specificity of the text-search algorithm alone (i.e., not in combination with manual review of patient records) was calculated [Table 3]. This was done by labeling all reports that were excluded through manual chart review as false positive reports (i.e., reports that were positively identified by the text-search algorithm but ended up not being associated with a rectal cancer resection). A total of 5588 reports were successfully identified through text-search from 284,032 total reports. 4550 false-positive reports were excluded through manual chart review, resulting in a specificity of
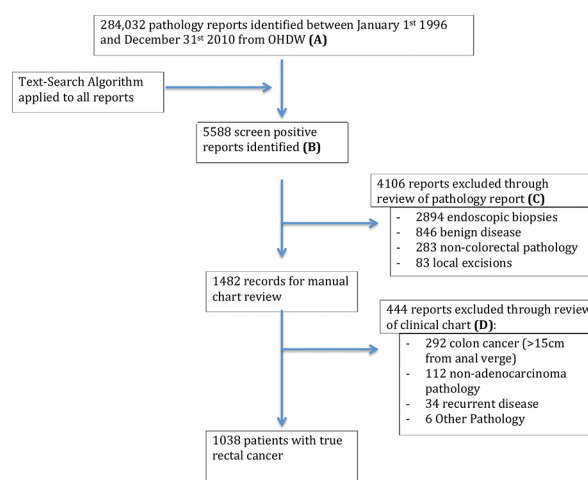

Figure 2: Flow diagram illustrating identification and validation of all rectal cancer resections at The Ottawa Hospital between January 1, 1996 and December 31, 2010. OHDW: Ottawa Hospital Data Warehouse

98.4%. This is essentially identical to the specificity of 98.6% calculated during the testing phase of the text-search algorithm [Table 2], and resulted in a PPV of 18.6% and a negative predictive value of 100%.

Discussion

In the present study, we created a cohort of patients who had a surgical resection for rectal cancer through a novel text-search approach of pathology reports. When measured against a cohort of 694 known rectal cancer resections with 2000 nonrectal cancer patients, this algorithm produced a sensitivity of 100% and specificity of 98.6%. When applied to all pathology reports

Table 3: 2×2 table demonstrating sensitivity, specificity, positive predictive value, and negative predictive value of a text-search algorithm applied to pathology reports for rectal cancer resections in Ottawa Hospital Data Warehouse (*n*=284,032)

Text search identified rectal cancer resections#	True rectal cancer resections*		Totals
	Positive	Negative	
Positive	A - 1038	B - 4550	5588
Negative	C - 0¶	D - 278,444	278,444
Totals	1038	282,994	284,032

Sensitivity=A/(A + C)=1038/1038=100%.
Specificity=D/(B + D)=278,444/282,994=98.4%.
PPV=A/(A + B)=1038/5588=18.6%.
NPV=D/(C + D)=278,444/278,444=100%.
*True rectal cancer resections refer to the validated cohort of rectal cancer patients created through manual chart review in the present study, #Text search identified rectal cancer resections refers to the rectal cancer resections identified through the text-search algorithm of pathology reports within the OHDW, ¶Assumes 100% sensitivity of text-search algorithm. PPV: Positive predictive value, OHDW: Ottawa Hospital Data Warehouse, NPV: Negative predictive value

within the OHDW, a total of 284,032 reports, specificity was maintained at 98.4%, yielding a PPV of 18.6%.

To date, no published data on the use of text-search methods to identify rectal cancer surgery cohorts specifically are available. However, our results compare favorably to the published literature on using text-search methodology for cohort creation of different patient populations. Nelson *et al*. developed a method to electronically search and categorize pathologic diagnoses of breast cancer patients based on text-search of pathology reports for biopsies and surgical resections and found 97.5% agreement with manual chart review.[17] Several other studies have found that the use of text-search methods to identify surgical complications and patients with ischemic heart disease have similar or better sensitivity and specificity when compared to methods that rely on diagnostic or procedural codes.[18,22] Despite its relative success, evidence suggests that text-search methods for cohort identification and dataset creation are underutilized.[23]

Several limitations exist for our study. First and foremost is the low pretest probability for rectal cancer in our study population. Only 1038 of 284,032 total pathology reports were for rectal cancer, yielding a pretest probability of just 0.36%. This very low disease prevalence explains the low PPV of only 18.6% for a text-search algorithm with a very high sensitivity and specificity (100% and 98.4%, respectively). Therefore, this method is not a viable option for cohort identification unless it is combined with manual chart review to validate the cohort. This may limit its usefulness when applied to larger, population-based databases for which access to individual patient records is not possible or feasible. One alternative would be to adjust the algorithm so that specificity is increased at the expense of sensitivity, but this would result in a greater number of missed patients. These results are consistent with the

notion that text-search methods usually increase sensitivity but lower specificity when compared to other methods of dataset creation using administrative data.[23]

A second limitation is the fact that our search was limited to pathology reports at a single institution, which may limit the generalizability of the results. For example, TOH introduced synoptic pathological reporting of rectal cancer since 2002. Synoptic reporting, with predetermined subject headings and repeating terminology, may lend itself to text-search methods with improved success. Similar methods applied to text documents that do not contain synoptic reporting may not fare as well. Similarly, the OHDW contains actual text reports for both pathology and radiology that are continually uploaded into the data warehouse. Without this, text-search methods would be impossible. Larger, population-based databases may or may not contain the actual text necessary to carry out this method of cohort creation, thereby limiting the generalizability of the method.

Related to this limitation is the lack of external validation of our text-search algorithm. Because the text-search terms utilized in the overall algorithm were only applied to pathology reports of a single institution, they were likely over-fitted to the data at this institution. To make this algorithm generalizable to research within other institutions and datasets, it should be tested on pathology reports generated outside TOH.

Finally, as previously mentioned, the definition of rectal cancer can vary, making differentiating it from sigmoid colon cancer sometimes difficult. For the purposes of this study, we used the widely accepted definition for rectal cancer of cancer that lies within 15 cm of the anal verge.[13] However, pathology reports may or may not use this definition, focusing their diagnosis on cancer of the colon instead. Such a subtle distinction may not be relevant for other diagnoses such cancer of the breast or pancreas, making text-search methods more relevant for these other diagnoses. Therefore, the text-search method can be limited depending on the type of patient one is trying to identify administrative data.

CONCLUSIONS

The present study created and validated a cohort of all rectal cancer resections performed at TOH over a 15-year period using a novel text-search method of pathology reports combined with manual chart review. The text-search algorithm developed yielded a very high sensitivity and specificity of 100% and 98.4%, respectively. This suggests text-search of pathology reports is a viable way to identify rectal cancer patients within limited datasets such as cancer registries. However, in our study, where the denominator was greater than 600,000 hospital admissions (of varied diagnoses) we could only produce a PPV of 18.6% despite a very high specificity. This means each report positively identified by the algorithm had an 18.6% probability of having a diagnosis of rectal cancer. Therefore, in the presence of a low pretest

probability, as is the case in large, population-based databases, text-search methods of pathology reports yield too many false positives despite a very high specificity. In these circumstances, text-search should be combined with a validation method, such as manual chart review, to be a viable approach. Alternatively, more complex text-search techniques such as natural language processing have shown potential for improved discrimination within clinical reports, which may become more relevant with the widespread adoption of electronic medical records.[24] Regardless, future research should look at the possibility of utilizing text-search as a tool for cohort identification in larger, population-based administrative datasets. Validation of such methods would improve the validity of administrative data research by ensuring accurate cohort identification.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

References

1. Virnig BA, McBean M. Administrative data for public health surveillance and planning. Annu Rev Public Health 2001;22:213-30.
2. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently used validated diagnostic or procedural codes. J Clin Epidemiol 2011;64:1054-9.
3. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. J Public Health Med 2001;23:205-11.
4. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. Med Care 2004;42:1066-72.
5. Hohl CM, Karpov A, Reddekopp L, Doyle-Waters M, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: A systematic review. J Am Med Inform Assoc 2014;21:547-57.
6. Prentice RL. Surrogate endpoints in clinical trials: Definition and operational criteria. Stat Med 1989;8:431-40.
7. Godwin JD 2nd, Brown CC. Some prognostic factors in survival of patients with cancer of the colon and rectum. J Chronic Dis 1975;28:441-54.
8. Wolmark N, Wieand HS, Rockette HE, Fisher B, Glass A, Lawrence W, et al. The prognostic significance of tumor location and bowel obstruction in dukes B and C colorectal cancer. Findings from the NSABP clinical trials. Ann Surg 1983;198:743-52.
9. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American joint committee on cancer sixth edition staging.

10. Edge SB, Byrd DR, Compton CC, Fritz AG, Greene FL, Trotti A, et al., editors. American Joint Committee on Cancer. Cancer Staging Manual. 7th ed. New York: Springer; 2010. p. 143.
11. Wong R, Berry S, Spithoff K, Simunovic M, Chan K, Agboola O, et al. Preoperative or Postoperative Therapy for the Management of Patients with Stage II or III Rectal Cancer: Guideline Recommendations; July, 2008. Available from: https://www.cancercare.on.ca/cms/One.aspx?portalId=1377&pageId=10207. [Last accessed on 2016 Dec 19].
12. Tjandra JJ, Kilkenny JW, Buie WD, Hyman N, Simmang C, Anthony T, et al. Practice parameters for the management of rectal cancer (revised). Dis Colon Rectum 2005;48:411-23.
13. van der Pas MH, Haglind E, Cuesta MA, Fürst A, Lacy AM, Hop WC, et al. Laparoscopic versus open surgery for rectal cancer (COLOR II): Short-term outcomes of a randomised, phase 3 trial. Lancet Oncol 2013;14:210-8.
14. Pilipshen SJ, Heilweil M, Quan SH, Sternberg SS, Enker WE. Patterns of pelvic recurrence following definitive resections of rectal cancer. Cancer 1984;53:1354-62.
15. Colorectal Cancer Collaborative Group. Adjuvant radiotherapy for rectal cancer: A systematic overview of 8,507 patients from 22 randomised trials. Lancet 2001;358:1291-304.
16. Guillou PJ, Quirke P, Thorpe H, Walker J, Jayne DG, Smith AM, et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): Multicentre, randomised controlled trial. Lancet 2005;365:1718-26.
17. Nelson HD, Weerasinghe R, Martel M, Bifulco C, Assur T, Elmore JG, et al. Development of an electronic breast pathology database in a community health system. J Pathol Inform 2014;5:26.
18. Ivers N, Pylypenko B, Tu K. Identifying patients with ischemic heart disease in an electronic medical record. J Prim Care Community Health 2011;2:49-53.
19. Ng KH, Ng DC, Cheung HY, Wong JC, Yau KK, Chung CC, et al. Laparoscopic resection for rectal cancers: Lessons learned from 579 cases. Ann Surg 2009;249:82-6.
20. Stevenson AR, Solomon MJ, Lumley JW, Hewett P, Clouston AD, Gebski VJ, et al. Effect of laparoscopic-assisted resection vs. open resection on pathological outcomes in rectal cancer: The ALaCaRT randomized clinical trial. JAMA 2015;314:1356-63.
21. van Walraven C, Wong J, Morant K, Jennings A, Jetty P, Forster AJ, et al. Incidence, follow-up, and outcomes of incidental abdominal aortic aneurysms. J Vasc Surg 2010;52:282-90.
22. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. JAMA 2011;306:848-55.
23. McKenzie K, Scott DA, Campbell MA, McClure RJ. The use of narrative text for injury surveillance research: A systematic review. Accid Anal Prev 2010;42:354-63.
24. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical text analysis and knowledge extraction system (cTAKES): Architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17:507-13.