

**Using Administrative Databases to Measure Surgical Quality for Rectal
Cancer at The Ottawa Hospital from 1996-2010**

Reilly Patrick Musselman

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements
for the Master's of Science degree in Epidemiology

School of Epidemiology

Faculty of Medicine

University of Ottawa

Table of Contents

Chapter	Page Number
<i>List of Tables...</i>	4
<i>List of Figures...</i>	5
<i>Legend of Frequently Used Acronyms and Symbols...</i>	6
<i>Acknowledgements...</i>	7
<i>Abstract...</i>	8
Chapter 1: Introduction	10
1.1 <i>Rectal Cancer...</i>	11
1.1.1 <i>Objective 1: Cohort Identification and Validation...</i>	13
1.2 <i>Surgical Quality and Rectal Cancer...</i>	15
1.2.1 <i>Surgical Quality of Rectal Cancer...</i>	17
1.3 <i>Administrative Databases...</i>	19
1.3.1 <i>Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer...</i>	20
1.4 <i>Summary of Objectives...</i>	21
Chapter 2: Methods	23
2.1 <i>Objective 1: Cohort Identification and Validation...</i>	24
2.2 <i>Objective 2: Surgical Quality of Rectal Cancer...</i>	27
2.3 <i>Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer...</i>	30
Chapter 3: Results	34
3.1 <i>Objective 1: Cohort Identification and Validation...</i>	35

3.2 Objective 2: Surgical Quality of Rectal Cancer...	37
3.3 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer...	40
Chapter 4: Discussion	42
4.1 Summary of Principle Findings...	43
4.2 Objective 1: Cohort Identification and Validation...	44
4.2.1 Discussion of Findings in Context of Current Literature...	44
4.2.2 Strengths and Limitations...	46
4.2.3 Conclusions and Future Considerations...	48
4.3 Objective 2: Surgical Quality of Rectal Cancer...	48
4.3.1 Increased Rectal Cancer Resections...	50
4.3.2 Lymph Node Retrieval...	52
4.3.3 Circumferential Resection Margin...	54
4.3.4 Distal Resection Margin and APR Rate...	56
4.3.5 Strengths and Limitations...	60
4.3.6 Conclusions and Future Considerations...	62
4.4 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer...	63
4.4.1 Discussion of Findings in Context of Current Literature...	63
4.4.2 Strengths and Limitations...	65
4.4.3 Conclusions Future Considerations...	68
References...	70

List of Tables

Table Caption	Page #
<i>Table 1. Inclusion and exclusion criteria used to define rectal cancer cohort.....</i>	83
<i>Table 2. First round results of text-search algorithm* development. Each iteration was performed on a population of 694 known rectal cancer resections performed at The Ottawa Hospital.....</i>	84
<i>Table 3. Second round results of text-search algorithm* development. 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, non-colorectal cancer cases.....</i>	85
<i>Table 4. 2x2 table demonstrating sensitivity, specificity, PPV and NPV of previously published coding algorithm for identifying rectal cancer resections in ICES when compared to a validation cohort (N=664,075).....</i>	86
<i>Table 5. Demographics for entire rectal cancer cohort created in Objective 1 (N=1038).</i>	87
<i>Table 6. Trends for surgical quality indicators for rectal cancer surgery at The Ottawa Hospital between January 1st, 2000 and December 31st, 2010 (N=972).....</i>	88
<i>Table 7. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome Circumferential Resection Margin (CRM)....</i>	89
<i>Table 8. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome Successful Lymph Node Retrieval.....</i>	90
<i>Table 9. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome Abdominal Perineal Resection.....</i>	91
<i>Table 10. 2x2 table demonstrating sensitivity, specificity, PPV and NPV of previously published coding algorithm for identifying rectal cancer resections in ICES when compared to a validation cohort (N=664,075).....</i>	92

List of Figures

Figure Caption	Page #
<i>Figure 1. Anatomical divisions of the colon and rectum</i>	93
<i>Figure 2. Graphical representation of mesorectal fascial planes followed during Total Mesorectal Excision (TME) (from Heald et al., 1982).....</i>	94
<i>Figure 3. Flow diagram depicting methodology used to complete Objective 1 (Identification and validation of rectal cancer cohort) Letters A-G represent specific steps referred to in text. The final cohort of rectal cancer patients was created from the combination of box H and I (Figure 4).....</i>	95
<i>Figure 4. Flow diagram depicting methodology used for manual chart review to confirm diagnosis of rectal cancer and identify relevant covariates.....</i>	96
<i>Figure 5. Flow diagram depicting methodology used to complete Objective 2 (assessment of surgical quality indicators).....</i>	97
<i>Figure 6. Flow diagram illustrating methodology used to identify rectal cancer resections through administrative databases in previously published studies (re-created with permission from the authors^{39,40}).....</i>	98
<i>Figure 7. Flow Diagram illustrating methods used to identify pathology reports of known rectal cancer resections at The Ottawa Hospital from 2002 to 2010 (*This group of reports was later used to develop a text-search algorithm for accurately identifying rectal cancer resections through administrative databases).....</i>	99
<i>Figure 8. Flow Diagram illustrating identification and validation of all rectal cancer resections at The Ottawa Hospital between January 1st, 1996 and December 31st, 2010.....</i>	100
<i>Figure 9. Total number of rectal cancer resections performed per year at The Ottawa Hospital between 1996 and 2010 (inclusive).....</i>	101
<i>Figure 10. Flow Diagram illustrating patient records from validation cohort that were successfully uploaded to ICES databases.....</i>	102

Legend of Frequently Used Acronyms and Symbols

<i>Acronym or Symbol</i>	<i>Definition</i>
<i>TOH</i>	<i>The Ottawa Hospital</i>
<i>TOH-CRC Registry</i>	<i>The Ottawa Hospital – Colorectal Cancer Registry</i>
<i>OHDW</i>	<i>Ottawa Hospital Data Warehouse</i>
<i>ICES</i>	<i>Institute for Clinical and Evaluative Sciences</i>
<i>CRM</i>	<i>Circumferential Resection Margin</i>
<i>DRM</i>	<i>Distal Resection Margin</i>
<i>APR</i>	<i>Abdominoperineal Resection</i>
<i>TME</i>	<i>Total Mesorectal Excision</i>
<i>MRI</i>	<i>Magnetic Resonance Imaging</i>

Acknowledgements

I would like to acknowledge the following contributions to this project:

Dr. Husein Moloo, Dr. Robin Boushey, and Dr. Rebecca Auer are general surgeons at The Ottawa Hospital who specialize in the surgical management of rectal cancer. Their clinical expertise and feedback was invaluable for both the planning of this project and the interpretation of results. In particular, I would like to thank Dr. Auer for providing access to data in the TOH-CRC registry.

Data analysts are required to facilitate access to data in both ICES and OHDW related databases. In this regard, I would like to acknowledge the following specific contributions: Josee Blackburn and Allison Jennings assisted with the extraction of pathology reports from OHDW; Deanna Rothwell and Jocelyn Tufts assisted in uploading the validation cohort created in our study onto ICES servers; Angela Raymond helped facilitate approval of the required Data Sharing Agreement allowing us to share data on ICES servers.

Finally, the help and mentorship provided by my supervisor, Dr. Carl van Walraven cannot be overstated. His expertise in planning and executing data collection and analysis throughout the entire project was invaluable. Also, as an ICES scientist, he was able to provide access to ICES servers, which was required for Objective #3 of this study. This access allowed us to simulate previously described methods for identifying rectal cancer resections in ICES databases. His expertise in SAS code techniques then allowed us to compare the resulting dataset with our gold-standard dataset for the validation portion of the study. Most of all, however, Dr. van Walraven's support and encouragement helped me to overcome the many obstacles encountered while trying to complete this project. I cannot thank him enough.

For BS.....See you soon!

Abstract

Purpose: The purpose of this thesis was threefold: 1) To explore the use of text-search methods for identifying rectal cancer patients in large datasets; 2) To examine temporal trends of surgical quality indicators for rectal cancer at a single, tertiary-care institution; 3) To validate the use of administrative codes for identifying rectal cancer patients in population-based datasets.

Methods: 1) A text-search algorithm was developed, validated, and applied to all pathology reports at The Ottawa Hospital (TOH) over a 15-year period. Positive records were confirmed through manual chart review, and a gold-standard cohort of all rectal cancer resections performed at TOH was created. 2) Univariate and multivariate analyses were performed to assess temporal trends and associated factors for four (4) key surgical quality indicators. 3) Previously published methods for identifying rectal cancer resections in population-based datasets were validated using the cohort of patients created in Objective 1 as a gold standard.

Results: 1) The text-search algorithm had a sensitivity and specificity of 100% and 98.4%, respectively. Because of low disease prevalence, positive predictive value (PPV) was 18.6%. 2) The proportion of resections with successful lymph node retrieval improved significantly over the course of the study period. No change was demonstrated for the remaining 3 surgical quality indicators. 3) Previously described methods that utilize procedure codes to identify rectal cancer resections in large administrative datasets had a sensitivity and specificity of 89.5% and 99.9%, respectively, with a PPV of 64.9%.

Conclusions: It is feasible to utilize both procedure codes and text-search methods to identify patients with surgical resections for rectal cancer in administrative datasets. However, these methods are at risk of being inaccurate and resulting cohorts should be validated. Creating large cohorts of rectal cancer patients can be useful for studying a variety of subjects, including surgical quality.

Chapter 1: Introduction

1.1. Rectal Cancer

Colorectal cancer is an important disease. It is the third most common cancer overall and it is the second most common cause of cancer deaths in Canada for both men and women.¹ Over the past several decades, however, it has become apparent that rectal cancer is a disease process that is distinct from cancer in the rest of the colon, with both treatment protocols and outcomes that differ significantly.

The primary reason for this difference relates to the anatomy of the colon and rectum (Figure 1). The colon consists of several anatomical divisions. It begins at the ileocecal valve, where the small intestine attaches to the first part of the colon called the cecum. The cecum then proceeds distally into the ascending, or right, colon. Continuing to proceed proximal to distal from the ascending colon leads to the hepatic flexure, transverse colon, splenic flexure, descending colon, sigmoid colon, rectum, and finally the anus. The ascending and descending colons are fixed by peritoneum to the right and left lateral abdominal walls, respectively, while the transverse and sigmoid colons are free from peritoneal attachments (Figure 1).

The key distinction when discussing colon versus rectal cancer is the anatomical transition from the sigmoid colon to the rectum. While the clinical definition of the transition from the sigmoid colon to the rectum is vague, the histopathological definition utilizes the fact that the rectum is, at least partially, covered by a fascial lining known as the mesorectum. The mesorectum is the peritoneal investment enveloping the rectum, containing its associated lymph

nodes and vascular supply.² In contrast, the sigmoid colon is covered only by visceral peritoneum.² The lower third of the rectum lies completely outside the abdominal cavity (i.e. it is 'extraperitoneal') and is completely covered by the mesorectum while the upper two thirds are only partially covered by the mesorectum.²

The anatomical difference between the colon and rectum translates into significantly worse outcomes for patients with rectal cancer compared to those with cancer anywhere else in the colon. Several multivariate logistic regression analyses have identified that a tumor located in the rectum is an independent predictor of increased local recurrence rates and decreased survival. For example, Godwin and Brown found that the 5-year survival for colon and rectal cancer was 32% and 25%, respectively, after adjusting for tumor stage, age, and patient gender.³ Similarly, data from the NSABP clinical trials demonstrated a 3-fold increase in the relative risk of treatment failure for rectal cancer compared to colon cancer.⁴ While treatment regimens for both colon and rectal cancer have improved since these data were reported, the American Joint Committee on Cancer persistently reports poorer survival rates for patients with rectal cancer compared to colon cancer after adjusting for disease stage.⁵⁻⁶

In addition to worse outcomes, the *treatment* of rectal cancer differs significantly from colon cancer. While the mainstay for both diseases is surgical resection, the treatment of rectal cancer also involves the use of neoadjuvant (i.e. pre-operative) and adjuvant (i.e. post-operative) radiation therapy, with or without concurrent chemotherapy. Currently, neoadjuvant chemoradiation is the

accepted standard of care for locally advanced rectal cancer. Locally advanced rectal cancer includes T3 tumors (i.e. those invading through the muscularis mucosa layer of the bowel wall into the outer serosal layer), T4 tumors (i.e. tumors invading through the entire wall of the rectum and/or invading surrounding structures within the pelvis), tumors with suspected lymph node involvement, or tumors that threaten the circumferential resection margin.⁷⁻⁸ Similarly, adjuvant radiation therapy is used to decrease recurrence rates for locally advanced rectal cancers.⁷⁻⁸ This approach differs significantly from the treatment algorithm for colon cancer, which does not include the use of radiation therapy (either pre or post-operatively) and only utilizes preoperative chemotherapy in the setting of distant metastases.⁹ There are also significant differences in surgical treatment between colonic and rectal cancers which will be discussed in more detail in Section 1.2 below.

1.1.1 Objective #1: Cohort Identification and Validation

It is clear that rectal cancer differs significantly from colon cancer with respect to both approach to treatment and disease outcomes. This suggests that clinical studies focusing on rectal cancer should be conducted independently from those focusing on colon cancer. Early randomized controlled trials would often include both populations while performing a subset analysis for the rectal cancer cohort.¹¹⁻¹² However, more recently investigators have recognized that questions regarding rectal cancer require their own focus, thus leading to large trials of rectal cancer that have excluded patients with cancer in the colon.¹³⁻¹⁴

Similarly, many observational studies that have included both colon and rectal cancer populations routinely perform subset analysis on the portion of the study population that has rectal cancer.¹⁴⁻¹⁵

However, differentiating cancer of the rectum from that of the colon for the purposes of clinical research can be challenging. This is because the definition of rectal cancer is vague and can vary from one study to another. Specific definitions of rectal cancer can include: 1. tumors that are within 15cm of the anal verge; 2. tumors that lie below the peritoneal reflection of the abdominal cavity (i.e. the division between the abdominal cavity and the pelvis); 3. Tumors that are amenable to radiation therapy; or simply 4. Tumors that are described as rectal cancer by the clinician. This varied definition is less of a concern in randomized controlled trials since definitions can be made *a priori* allowing for a homogenous study population. However, the vagaries of defining rectal cancer – even with primary data – is more problematic in retrospective studies. This error would be amplified in retrospective studies based on health administrative data wherein cases are identified using diagnostic codes. In this situation, error associated with classifying rectal cancer is added to error from assigning a diagnostic code based on medical record review.

Statement of Objective #1: Develop a text-search algorithm to identify all pathology reports of patients with surgical resections for rectal cancer at The Ottawa Hospital from 1996-2010 for the purpose of creating and validating a patient cohort consisting of all rectal cancer resections at The Ottawa Hospital over this time period.

Therefore, the first objective of the current project was to create a cohort of all patients who underwent a surgical resection for rectal cancer at a single, tertiary-care institution over a 15-year period. The aim was to accomplish this by searching pathology reports and validate the cohort through manual chart review. Creating such a population would allow for investigation and analysis on a large cohort of rectal cancer resections with assurance of the accuracy of the population.

1.2 Surgical Quality and Rectal Cancer

Several theories have been postulated to explain the worse clinical outcomes seen in rectal cancer when compared to cancer in the rest of the colon. One study suggested that tumor cell DNA changes accounted for more aggressive behaviour of rectal cancers, but this hypothesis was eventually debunked.¹⁶ Other studies have also suggested that rectal tumors are biologically more aggressive than tumors of the colon, but these theories have yet to be proven.^{4,17}

The more likely explanation is that the worse outcomes are related to the actual anatomy itself. Located in the pelvis, rectal tumors are much more likely to invade surrounding structures due to their close proximity to these surrounding tissues. Such structures include the prostate, vagina, sacrum, bladder and pelvic side-wall.² Similarly, the narrow confines of the pelvis make complete surgical excision of a tumor much more technically challenging, increasing the possibility of leaving tumor cells behind.

It has therefore been postulated that anatomy, and not tumor biology, is the driving force behind the poor outcomes for rectal cancer.¹⁶ This theory was supported by Quirke and colleagues, who discovered that high recurrence rates in rectal cancer compared to colon cancer were the result of circumferential tumor spread into this mesorectum.¹⁸⁻¹⁹ Specifically, they found that positive circumferential margins (i.e. leaving tumor cells behind after a surgical resection) could be used to predict local recurrence with a sensitivity and specificity of 92% and 95%, respectively.¹⁸⁻¹⁹

This discovery led to significant changes in the way rectal cancer is treated. From a surgical perspective, Heald and colleagues pioneered the surgical technique known as the Total Mesorectal Excision (TME), an operation that uses sharp dissection to remove the entire mesorectum (Figure 2).²⁰ This technique was developed to reduce the rate of positive circumferential resection margins (CRM) following surgery, thereby decreasing local recurrence rates and improving overall survival. The CRM is defined as the distance between the rectal tumor and the nearest margin in the direction that moves radially from the lumen of the rectum towards the rectal wall. Since its introduction, the TME has become the standard of care for patients with surgically resectable rectal cancer. Results from randomized controlled trials have clearly demonstrated a decreased recurrence rate and improved survival when TME was used instead of traditional techniques.²¹ Similarly, the introduction and training of surgeons on TME technique have also improved long term outcomes.²² Specifically, a comparison of trials measuring outcomes following rectal cancer surgery before and after the

implementation of TME as the standard surgical approach reported a decrease in local recurrence rates from 16% to 9%.²²

Total Mesorectal Excision adheres to the principle of complete surgical excision of the cancer while maintaining negative resection margins, also known as an R0 resection. While the TME is designed to reduce the rate of positive CRM, the principle of an R0 resection should be maintained for all resection margins. For example, it has been demonstrated that even with an adequate TME and appropriate adjuvant chemoradiation therapy, surgical excision with a positive distal margin (defined as the distance between the tumor and the nearest resection margin moving in the downstream direction towards the anus) will have significantly higher rates of local recurrence and mortality.²³⁻²⁴ Similarly, the number of lymph nodes retrieved and examined in surgical resection has been shown to have prognostic significance for both colon and rectal cancer.²⁶⁻²⁷ While both the surgeon and the pathologist play a role in the number of lymph nodes identified and examined, lymph node retrieval has also been associated with CRM and overall adequacy of resection, indicating that it is an important quality indicator for surgical resection of rectal cancer.²⁷

1.2.1 Objective #2: Surgical Quality of Rectal Cancer

It is clear in the literature that there is a direct link between the quality of surgical resection for rectal cancer and improvement in significant patient outcomes such as local recurrence and overall survival.

However, several studies have indicated that outcomes for rectal cancer vary extensively on a population level. Specifically, local recurrence rates ranging from 4-19% and relative 5-year survival ranging from 32-64%.²⁸⁻²⁹ While some of this variation has been attributed to different staging techniques, variation in surgical practice has also been implicated.³⁰ In fact, Porter and colleagues found that a lack of subspecialty training in colorectal surgery and low surgeon case-volumes were independent predictors of increased local recurrence rates and decreased survival in patients with rectal cancer.³¹ This is supported by a multitude of data that correlates surgical quality with case volume in rectal cancer surgery.³²⁻³⁴

These data imply that, not only is the quality of surgical resection for rectal cancer directly related to patient outcomes, but there appears to be significant variation in said surgical quality. The Ottawa Hospital (TOH) has introduced a number of interventions over the past 15 years to improve quality of care for patients with colon and rectal cancer. These include the implementation of a Communities of Practice (CoP) model of patient care that centralizes treatment of rectal cancer at TOH to surgeons with subspecialty training in colorectal surgery or surgical oncology. It also includes educational initiatives for surgeons, pathologists and oncologists regarding important quality markers for the treatment of rectal cancer.³⁵⁻³⁶ The impact of these interventions on surgical quality for rectal cancer have, to date, not been assessed.

Statement of Objective #2: Use data from the patient cohort created in Objective #1 to measure trends in surgical quality of rectal cancers at The Ottawa Hospital using recognized quality indicators.

Therefore, objective #2 will report on trends of surgical quality for rectal cancer at TOH using recognized surgical quality indicators. Such an audit would provide objective data that could potentially be used for assessment of interventions aimed at improving quality of rectal cancer surgery at our institution. It would also be a useful method to identify areas that have improved and areas that need to be improved in rectal cancer surgery at The Ottawa Hospital.

1.3 Administrative Databases

The results of Objectives #1 and #2 will yield an accurate cohort of patients who have undergone surgery for rectal cancer at a single institution with an assessment of their surgical quality. Creating the same cohort on a larger, population-based scale can be more problematic since manually reviewing tens of thousands of clinical records is not practically feasible. One manner in which researchers could assemble and analyze large quantities of data is through of the use of administrative databases. Administrative data are defined as “electronic data routinely collected and abstracted for administrative purposes”.³⁷ Advantages of utilizing such datasets include the ability to conduct research on a population scale, relatively easy access to large quantities of information, and the fact that most databases are routinely collected allowing for access to the most up to date population level data.³⁷

Traditional administrative databases rely on diagnostic and procedural codes to identify patient cohorts. The accuracy of these codes for the disease or procedure they are supposed to represent is often less than perfect. Van Walraven and colleagues demonstrated that, while diagnostic and procedure codes from administrative datasets are commonly used to create patient cohorts, these codes are rarely validated; this calls into question the accuracy of the cohorts created using these codes.³⁸ This landmark study also concluded that the probability that the codes used in administrative data studies actually identified the disease or procedure intended was often less than 50%. Inaccurate codes will result in poorly defined patient cohorts in which some of the included patients will actually *not* have the disease of interest and many with the disease will be excluded. This would obviously call into question the validity of the results from such studies.

1.3.1 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer

Several studies have reported on outcomes for colorectal cancer surgery on a population scale using administrative data.³⁹⁻⁴¹ The majority of which utilized non-validated procedure codes to identify their patient cohort. Traditional procedure codes for rectal surgery include those for anterior resections, low anterior resections, and abdominal perineal resections. In addition to the inaccuracy of procedure codes in general, it is especially difficult to differentiate cancer of the rectum from cancer of the sigmoid colon by simply using procedure

codes. That is, even if the procedure codes are accurate, they do not themselves pinpoint the location of the tumor and are often used interchangeably for surgical resection of both sigmoid and rectal cancers.

Statement of Objective #3: Determine the accuracy of currently described methods for identifying rectal cancer resections using administrative databases by comparing the population derived using these methods to the gold-standard population created in Objective #1.

Despite these challenges, the methodology used to identify rectal cancer resections with administrative data has, to date, not been validated. It is therefore difficult to know whether or not the methods used to identify cases of rectal cancers accurately identified this cohort. The third objective of this project was therefore to perform a validation study that compares a cohort of patients created using previously described methods in the scientific literature with the gold-standard population created in Objective #1 of the current study.

1.4 Summary of Objectives

The present thesis project attempted to identify every patient with rectal cancer having a surgical resection at The Ottawa Hospital from 1996-2010 and measure the quality of their surgery. Using this cohort as a gold standard, a validation study was performed to measure the accuracy of previously described methods for identifying rectal cancer resections using population scale administrative databases. Such a study will inform future studies as to whether or

not these methods require improvement so rectal cancer populations can be reliably studied on a population scale.

Statement of Objectives:

- 1) ***Develop a text-search algorithm to identify all pathology reports of patients with surgical resections for rectal cancer at The Ottawa Hospital from 1996-2010 for the purpose of creating and validating a patient cohort consisting of all rectal cancer resections at The Ottawa Hospital over this time period.***

- 2) ***Use data from the patient cohort created in Objective #1 to measure trends in surgical quality rectal cancer at The Ottawa Hospital using recognized quality indicators.***

- 3) ***Evaluate currently described methods for identifying rectal cancer resections using administrative databases by comparing the population derived using these methods to the gold-standard population created in Objective #1.***

Chapter 2: Methods

2.1 Objective 1: Cohort Identification and Validation

All patients who had a surgical resection for rectal cancer at The Ottawa Hospital from 1996-2010 were identified using the Ottawa Hospital Data Warehouse (OHDW) using the following algorithm (Figure 3):

1. Patients who had a surgical resection for Colon and Rectal Cancer (CRC) between 2002 and 2010 at The Ottawa Hospital were identified using International Classification of Disease, version 10 (ICD-10) procedure codes (ICD-10 codes used in this process are summarized in Appendix 1). These patients were recorded in a retrospectively maintained registry by a surgical research group at the Ottawa Hospital Research Institute (OHRI) called the TOH-CRC registry. With permission from our research ethics board, patient identification and surgery dates of all rectal cancer resections were extracted from the TOH-CRC registry (Figure 3-A). The medical records for each of these encounters were reviewed. Patients meeting inclusion criteria for the present study were included (Table 1). This TOH-CRC dataset was then uploaded into the OHDW, wherein the health record numbers was encrypted to permit linkage with other patient-specific records within the OHDW (Figure 3-B).
2. The pathology reports for all of the resections in the TOH-CRC dataset were retrieved in the OHDW by linking the unique patient identifiers to their pathology reports in the OHDW. These reports were manually reviewed to

identify key clauses and phrases that would identify patients with rectal cancer (Figure 3-C). Using these phrases, a text-search algorithm was developed using the *%textMINER* SAS macro. This landmark SAS code macro, developed in Ottawa, allows the user to develop search algorithms for textual data and has been used in a previous MSc Epidemiology thesis project. This macro was used to identify a text search algorithm that was at least 99% sensitive for all rectal cancer surgery patients in the CRC database (Figure 3-D).

3. A random sample of 1000 pathology reports that did not have rectal cancer specimens, along with a sample of 1000 pathology reports of colon cancer specimens identified through the TOH-CRC database, were retrieved and added to the cohort identified in the previous step (Figure 3-E). Each of these non-rectal cancer pathology reports was reviewed prior to insertion into the algorithm to confirm appropriate designation of disease. This resulted in a sample of 2694 pathology reports for which the presence or absence of rectal cancer was known (i.e. 1000 colon cancer reports; 1000 non-rectal, non-colon cancer reports; and 694 known rectal cancer reports from the TOH-CRC registry). The text search algorithm that had been developed in step 3-C was applied to these 2694 reports to measure the sensitivity and specificity of the algorithm for rectal cancer on pathology reports. The text search algorithm was then modified to maximize its sensitivity and specificity. Once complete,

the operating characteristics of the text-search algorithm for rectal cancer resections were determined and the final algorithm created (Figure 3-F).

4. The final pathology report text search algorithm was then applied to *all* pathology reports recorded in the OHDW (Figure 3-G). All screen-positive pathology reports identified by the algorithm were linked to the TOH-CRC registry to identify any cases in the latter that were missed. Missed cases were added to the cohort.

The pathology reports of all cases in the final cohort were manually reviewed – along with their operative notes, clinic notes, radiology reports and hospitalization discharge summaries in vOacis – and the criteria described in Table 1 were applied to identify all study-eligible rectal cancer resections performed at The Ottawa Hospital between 1996 and 2010 (Figure 4-A). For the purposes of the present study, rectal cancer was defined as an adenocarcinoma of the rectum with the distal margin of the tumor lying within 15cm of the anal verge, which is consistent with the definition used in major clinical trials for the disease.⁴³⁻⁴⁴ Measurements based on rigid sigmoidoscopy were used preferentially. If such measurements were not readily available in the notes, measurements from the anal verge were based on, in order of preference, flexible endoscopy, MRI, and digital rectal exam. If no measurement could be found in any of the notes, determination of rectal cancer was made based on best clinical judgement using the information available in the chart.

If the operative note and discharge summary for the hospitalization was unavailable on the hospital's electronic medical record, the paper-based medical record of each patient's hospitalization was retrieved and the criteria listed in Table 1 was applied (Figure 4-B). If this hospital record was unavailable or inadequate, clinic notes from the Cancer Assessment Center were reviewed (Figure 4-C) to apply the criteria in Table 1. All patient records identified with the above algorithm that met inclusion criteria were included in the final cohort of patients who received a surgical resection for rectal cancer at The Ottawa Hospital between January 1st, 1996 and December 31st, 2010.

2.2 Objective 2: Surgical Quality of Rectal Cancer

Once the cohort was identified and validated through manual chart review, surgical quality was assessed. Methods used to complete this objective are summarized in Figure 5. As discussed in Section 1.2 of the introduction, four surgical related factors have been associated with improved outcomes in rectal cancer and were used as surgical quality indicators in the present study. These indicators were:

- 1) Circumferential Resection Margins (CRM): This has become an accepted indicator of the quality of surgical resection in patients with rectal cancer and has been shown to be an independent predictor of local recurrence and disease free survival in a follow-up study of clinical trial participants.¹⁸⁻
¹⁹ A successful outcome for this variable was defined as a negative CRM reported on the pathology report.

- 2) Abdominoperineal Resections (APRs): This was also identified as being an independent predictor of local recurrence and disease-free survival when compared to anterior and low anterior resections.²⁷ Resections performed using methods other than and APR was defined a successful outcome.
- 3) The number of lymph nodes retrieved and examined: This statistic has been shown to be of prognostic significance in patients with both colon and rectal cancer.²⁵⁻²⁶ Specifically, inadequate number of lymph nodes retrieved at the operation is known to negatively impact survival, likely by incorrectly “under-staging” patients. While both the surgeon and the pathologist play a role in the number of lymph nodes identified and examined, lymph node retrieval has also been associated with CRM and overall adequacy of resection,²⁷ indicating that it is an important quality indicator for surgical resection of rectal cancer. Although the minimum number of nodes to be resected is still somewhat controversial, current guidelines from Cancer Care Ontario support a minimum of 12 lymph nodes per surgical resection for adequate staging of colon and rectal cancer.⁴² For this quality indicator, a successful outcome was defined as 12 or more lymph nodes reported on the pathology report.
- 4) Distal resection margin (DRM): Like CRM, distal resection margin has been identified as an important prognostic indicator for patients undergoing a surgical resection for rectal cancer. Specifically, surgical excision with a positive DRM will have significantly higher rates of local

recurrence and mortality²³⁻²⁴. Unlike some prognostic factors, achieving a negative resection margin is largely under the direct control of the surgeon, and is therefore universally recognized as an indicator of the quality of the resection. A successful outcome for this variable was defined as a negative DRM reported on the pathology report.

These data were retrieved by manual review of each patient's pathology reports. Similarly, important covariates were also identified including: patient age; patient gender; medical comorbidities; cancer stage; and use of pre-operative chemoradiation. These covariates were identified through manual chart review using methods outlined in Figure 4.

Results were presented for 1-year time segments. Statistical independence in univariate analysis was determined using a Cochran-Mantel-Haenszel test to detect trends in binomial variables. Separate multivariable logistic regression models adjusting for key covariates were created to measure the association between procedure year and the odds of a successful outcome for each surgical quality indicator. Surgical year was introduced as a continuous variable and the outcome was the odds ratio for successful outcome of each of the four surgical quality indicators, all of which were binomial variables. Covariates used for adjustment included: patient age and gender; laparoscopic versus open surgery; emergency versus elective surgery; whether or not a multivisceral resection was performed; pathologic T-stage; pathologic N-stage; distance of the tumor from the anal verge; and the use of neoadjuvant chemoradiation treatment.

All predictor variables for the model were retrieved from manual review of the patient's hospital records. Final decision as to which predictor variables included in the model was made in conjunction with clinical experts. Univariate analysis comparing two levels of the main outcome for each predictor variables was performed. Multivariate logistic regression was then performed using both backward and stepwise selection as long as a minimum of 5 positive outcome events were identified for each variable in the model. Possible interactions between predictor variables were listed a priori and tested in the model. These included interactions between: age and emergency surgery; gender and multivisceral resection; and laparoscopic surgery and distance from the anal verge. Statistically significant interactions remained in the model. Missing data for any of the model variables resulted in removal of that patient's observations from the model entirely. The likelihood ratio test, Hosmer-Lemshow test and c-statistic were used to assess model goodness of fit. The model with the highest predictive value based on the c-statistic was used as the final model.

2.3 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer

The purpose of this objective was to evaluate the accuracy of an algorithm used in previously published studies to identify a cohort of patients who underwent a surgical resection for rectal cancer using health administrative data.³⁹⁻⁴⁰ The algorithm used procedure codes to identify surgical resections for rectal cancer using administrative data and was published in a peer-reviewed medical journal.⁴⁰ The algorithm utilized in this study was obtained with

permission from the authors and is represented in Figure 6. Briefly, this algorithm identified all patients in the province of Ontario between April 1st, 2002 and December 31st, 2010 with a diagnosis of colorectal cancer in the Ontario Cancer Registry (OCR). The OCR is a continually updated registry that provides information for all newly diagnosed solid-organ cancers in the province of Ontario and is under the umbrella of Cancer Care Ontario. ¹ Using the patient identifier and date of cancer diagnosis, these patients were linked to the Canadian Institute for Health Information Discharge Abstract Database (CIHI-DAD) to identify records for the hospitalization at which the pathology sample was extracted. International Classification of Disease, 10th edition (ICD-10) procedure codes were then used to identify those patients who had an anterior resection, low anterior resection, or abdominal-perineal resection (Appendix 1). The resultant cohort – which included all patients with rectal cancer in OCR who had a code for rectal cancer resection within 1-year of the pathology date – was thought to represent all patients in the province of Ontario who had a surgical resection for rectal cancer between April 1st, 2002 and December 31st, 2010. This study was performed through databases available at the Institute for Clinical and Evaluative Sciences (ICES). Databases were linked and analyzed using SAS statistical software, version 9.2 (SAS Inc., Cary, North Carolina).

For the purpose of the present study's validation, the above algorithm was reproduced limiting the cohort to all resections performed only at The Ottawa Hospital between April 1st, 2002 and December 31st, 2010. Access to relevant databases was obtained through permission of the ICES cancer group in Toronto,

ON. SAS code and ICD-10 procedure codes were obtained from the authors and reproduced at ICES Ottawa location. The SAS code used was obtained directly from the authors of the peer-reviewed study and was run in an identical fashion, with the only modification made for limiting the search to TOH patients. The final cohort was identified and labeled as the *study cohort*.

The validated cohort identified in objective #1 of the present study, heretofore referred to as the *validation cohort*, was uploaded onto the ICES server with permission from ICES@uOttawa. Since the validation cohort included all TOH patients between April 1st, 2002 and December 31st, 2010 with rectal cancer, patients who were not in this cohort definitely did *not* have rectal cancer.

The validation cohort was then linked to the *study cohort* using a common, unique identifier. With the *validation cohort* considered to be the gold standard population, sensitivity and specificity of the study cohort was calculated by determining the false negative and false positive rates of the algorithm utilized in the peer-reviewed study. A positive predictive value (PPV) was then calculated to determine the probability that someone identified as having a surgical resection for rectal cancer through the peer-reviewed study's algorithm actually had that procedure.

Ethics Approval

Ethics approval was obtained through the Ottawa Hospital Research Institute Research Ethics Board (OHRI REB) for access to both administrative databases contained within the OHDW and medical records.

Statistical Analysis

Objective #1 required calculation of sensitivity and specificity of the text search algorithm, as well as positive and negative predictive values. Analysis of outcomes reported in Objective #2 consist of reporting trends of surgical quality and comparing outcomes over the 15-year study period. Statistical independence was determined using a Mantel-Haenszel test for trend. Statistical analysis in Objective 3 consisted of the calculation of sensitivity, specificity, positive and negative predictive value, and likelihood ratio for the cohort identification algorithm used in previously published articles. All statistics will be compiled and analyzed using SAS statistical software, version 9.2 (SAS Inc., Cary, North Carolina).

Chapter 3: Results

3.1 Objective 1: Cohort Identification and Validation

720 patients from the TOH-CRC registry were identified as having a surgical resection for rectal cancer. After manual review of patient records, 694 met inclusion criteria for the present study (Figure 7). 724 pathology reports were identified for the 694 patient encounters, 30 of which were excluded as addendum reports (Figure 7). All of the included pathology reports were retrieved from the OHDW and were manually reviewed to identify key words and phrases from each. Multiple iterations of the text search algorithm were then attempted to adequately identify known rectal cancer resections from pathology reports. A total of 7 iterations of the text-search algorithm were applied to these 694 pathology reports, with each one utilizing different key words and phrases identified from the manual review. The goal of each iteration was to maintain high sensitivity while keeping the search terms as specific as possible. Sensitivity for each iteration presented in Table 2, with the best sensitivity score being 100%. Because this first round of testing was only performed on 694 pathology reports of known rectal cancer specimens, there were no negative reports to exclude, and therefore specificity could not be measured. The third iteration was chosen to proceed to the next round of testing because it maintained 100% sensitivity while using search terms most specific to rectal cancer resections (Table 2).

This text-search algorithm then was applied to a second population of pathology reports that contained the 694 known rectal cancers, 1000 known colon cancer resections, and 1000 randomly selected non-colon, non-rectal cancer patients. A total of 5 iterations of the text-search algorithm were applied

and the results are presented in Table 3. Different iterations were tested in order to find the combination of words and phrases that maintained 100% sensitivity while maximizing specificity. The final text-search algorithm produced a screening test that was 100% sensitive and 98.6% specific.

The final text-search algorithm was then used to identify all rectal cancer resections performed at The Ottawa Hospital between January 1st, 1996 and December 31st, 2010 (Figure 8). All pathology reports during this period were extracted from the OHDW, totaling 284,032 reports (Figure 8-A). After applying the final text-search algorithm to these reports, a total of 5588 screen positive pathology reports were identified (Figure 8-B). Manual chart review was then used to create the final cohort. Of the 5588 screen-positive reports identified from the text-search of all TOH pathology reports between January 1st, 1996 and December 31st, 2010, 4106 were excluded on the first-pass reading of the pathology reports themselves, leaving 1,482 for manual chart review of medical records, including operative reports, clinic notes, and radiology reports (Figure 8-C). This process excluded another 444 patients, leaving a total of 1038 patients who met inclusion criteria of the present study (Figure 8-D). When cross-referenced with the 694 patients initially identified from the CRC database, there were no missing records that needed to be added to the cohort.

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. This was done by labelling 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 98.4% (See Table 4 for calculations). This is essentially identical to the specificity of 98.6% calculated during the testing phase of the text-search algorithm (Table 3), and resulted in a positive predictive value (PPV) of 18.6% and a negative predictive value (NPV) of 100%.

3.2 Objective 2: Surgical Quality of Rectal Cancer

A total of 1038 rectal cancer resections were identified for analysis between January 1st, 1996 and December 31st 2010. Cohort demographics are presented in Table 5. Patients were middle-aged, mostly male, were equally distributed between stages 1 through 3, were predominantly done using an open approach, and greater than half had a low anterior resection (LAR),

The number of rectal cancer resections performed at The Ottawa Hospital increased steadily from 15 in 1996 to 105 in 2010 (Figure 9) and appeared to plateau after 2004. In addition to the number of resections performed, statistically significant increased trends were identified for the proportion of surgery performed laparoscopically ($p < 0.001$), the number of multivisceral resections ($p = 0.04$) and the proportion of patients undergoing neoadjuvant chemoradiation therapy ($p < 0.001$).

Table 6 summarizes the trends for surgical quality indicators over the course of for rectal cancer resections between January 1st, 2000 and December 31st, 2010. Data between 1996 and 2000 were excluded from the analysis due to a lack of reliable pathological reporting on several relevant quality indicators including circumferential resection margin and lymph node retrieval. Univariate analysis demonstrated a significant increase in the proportion of resections that successfully retrieved at least 12 lymph nodes ($p < 0.001$). No significant increase was demonstrated for the number of positive circumferential resection margins ($p = 0.24$), positive distal resection margins ($p = 0.97$) or the proportion of Abdominoperineal Resections ($p = 0.18$).

Multivariable logistic regression identified significant associations between predictor variables and each surgical quality outcome. The results are summarized in Tables 7 to 9. Positive circumferential resection margin (CRM) ranged from 2.9% in 2010 to 22.2% in 2001, with a mean CRM positivity rate of 6.8%. Positive CRM (Table 7) was significantly associated with multivisceral resections (OR=2.59, 95% CI = 1.24 – 5.40), emergency surgery (OR=10.11, 95% CI = 2.98 – 34.33), and distance of the tumor from the anal verge (OR = 0.88, 95% CI = 0.82 – 0.96). Also, compared to stage 1 tumors, higher stage tumors were significantly associated with a positive CRM (Stage 2 OR = 28.39, 95% CI = 3.73 – 215.95; Stage 3 OR = 31.01, 95% CI = 4.16-231.13; Stage 4 OR = 111.37, 95% CI = 10.00 - >999.99).

The proportion of resections with a successful lymph node retrieval ranged from 30.2% in 2000 to 87.6% in 2010 with an overall mean of 61.0%. Successful

lymph node retrieval (Table 8) was significantly less following the use of neoadjuvant chemoradiation (OR = 0.58, 95% CI = 0.43-0.80) but was significantly more likely as distance from the anal verge increased (OR = 1.04, 95% CI = 1.00 – 1.08), and with higher tumor stage (Stage 2 OR = 1.53, 95% CI = 1.04 – 2.24; Stage 3 OR = 1.71, 95% CI = 1.20-2.43; Stage 4 OR = 1.10, 95% CI = 0.40 – 3.00).

The proportion of cases with a positive distal resection margin (DRM) ranged from 0-2.6% with a mean of 0.6%. A logistic regression analysis was not performed due to the small number of events of a positive DRM. A stable logistic regression model requires 5-10 events per variable used in the model. Only 6 positive DRMs were identified over the 11-year study period with no events occurring in 7 of the 11 years, making it impossible to create a stable regression model.

The proportion of rectal cancer cases performed as abdominal perineal resections ranged from 26.1% in 2004 to 39.4% in 2000 with an overall mean of 33.7%. The odds of having an APR (Table 9) significantly increased in patients having a multivisceral resection (OR = 2.48, 95% CI = 1.01 – 6.08) and significantly decreased for tumors further away from the anal verge (OR = 0.45, 95% CI = 0.41 – 0.50).

The variable “year of surgery” was used in each model to evaluate changes of each quality indicator over time. Of the four quality indicator variables, only successful lymph node retrieval was significantly associated with year of surgery (OR = 1.27, 95% CI 1.21-1.33; $p < 0.01$). Circumferential resection

margin rates (OR = 1.00, 95% CI = 0.92 – 1.09) and proportion of APRs (OR = 0.97, 95% CI = 0.91 – 1.04) were not significantly associated with year of surgery in either respective multivariate regression model.

3.3 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer

886 records from the validation cohort had a surgery date between April 1st, 2002 and December 31st, 2010 (inclusive). These records were uploaded to the Institute for Clinical and Evaluative Sciences (ICES) servers. 821 of the 886 records (92.7%) were successfully uploaded to ICES, with 65 records excluded because of missing ICES key numbers (IKN), missing or invalid admission dates, or missing records in the Discharge Abstract Database (DAD) (Figure 10).

A total of 664,075 records were identified through the DAD in ICES for patients 18 years or older who were admitted to The Ottawa Hospital (TOH) between April 1st, 2002 and December 31st, 2010 (the same time period of the validation cohort). 22,180 of these patients (3.34%) had a diagnostic code for colorectal cancer in the Ontario Cancer Registry between one year prior to, or up to 14 days after, their admission date. 1,131 of these patients (5.1%) had at least one CCI procedure code in a DAD record that matched those listed in Appendix 1. This resulted in a study cohort of 1,131 records of rectal cancer resections identified through ICES datasets using previously described methods.

When compared to the validation cohort, sensitivity and specificity of the study cohort were 89.5% and 99.9%, respectively. However, because of the very low prevalence of true rectal cancer resections in this population (only 0.124% of

admissions), the positive predictive value (PPV) was only 64.9% (Table 10). This means that less than two thirds of patients identified with the codes in Appendix 1 truly had rectal cancer.

Chapter 4: Discussion

4.1 Summary of Principle Findings

The present study was divided into 3 main objectives. The principle findings for each objective are summarized below:

Objective 1 successfully created and validated a retrospective cohort of all rectal cancer resections performed at The Ottawa Hospital (TOH) between January 1st, 1996 and December 31st, 2010. This was achieved through a novel text-search algorithm of all pathology reports at TOH available through the Ottawa Hospital Data Warehouse (OHDW) followed by manual chart review. This process achieved a sensitivity of identifying cases of rectal cancer resections of 100% and a specificity of 96.5%.

Objective 2 measured four primary surgical quality indicators for rectal cancer surgery at The Ottawa Hospital between January 1st, 2000 and December 31st, 2010. Over the course of the study period, the proportion of resections with a successful lymph node retrieval increased significantly in both univariate and multivariate analyses. There was no significant change, however, in the other three quality indicators (rate of positive circumferential resection margin [CRM], positive distal resection margin [DRM], or the proportion of abdominal perineal resections [APRs]).

Objective 3 evaluated previously described methods that utilized procedure and diagnostic codes to identify patients who have had a rectal cancer resection through provincial level, population databases. When compared to the gold

standard cohort created in Objective #1 of the present study, these methods had a sensitivity and specificity of 89.5% and 99.9%, respectively, but a PPV of only 64.9%.

4.2 Objective 1: Cohort Identification and Validation

4.2.1 Discussion of Findings in Context of Current Literature:

Administrative data have become a common resource for clinical researchers. The use of administrative databases for healthcare related research publications has risen steadily over the past 2 decades, with results that often lead to important conclusions.⁴⁵ The principle method used to identify patients for dataset creation is through the use of codes related to the healthcare encounter. Examples include diagnosis codes such as the ICD-10 (International Classification of Disease, version 10), billing codes such as OHIP (Ontario Health Insurance Plan), and procedure codes such as CCI (Canadian Classification of health Interventions).

Several problems related to the use of healthcare related codes to identify patients through administrative datasets have arisen. First, a wide variability of the accuracy of administrative data codes has been reported.⁴⁶⁻⁴⁷ In addition to code accuracy, Hohl et al (2013) demonstrated a high degree of variability between studies in the numbers and types of codes used to identify the same population.⁴⁸ Finally, very few studies actually validate the codes they use to

create datasets for analysis.³⁸ As a result, novel ways to create reliable patient cohorts should be explored.

In the present study, we created a cohort of patients who had a surgical resection for rectal cancer via a novel text-search approach of pathology reports. When measured against a cohort of 694 known rectal cancer resections with 2000 non-rectal cancer patients, this algorithm produced a sensitivity of 100% and specificity of 98.6%. When applied to all pathology reports within the Ottawa Hospital Data Warehouse (OHDW), a total of 284,032 reports, specificity was maintained at 98.4%, yielding a positive likelihood ratio of 62.5.

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 favourably to the published literature on using text-search methodology for cohort creation of different patient populations. Nelson et al (2014) 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.⁴⁹ 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.⁵⁰⁻⁵¹ Despite its relative success, evidence suggests that text-search methods for cohort identification and dataset creation are underutilized.⁵²

4.2.2 Strengths and Limitations

Several limitations exist for our study. First and foremost is the low pre-test probability for rectal cancer in our study population. Only 1038 of 284,032 total pathology reports were for rectal cancer, yielding a pre-test probability of just 0.36%. This resulted in a positive predictive value of only 18.6%. 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 such as ICES (Institute for Clinical and Evaluative Sciences) or SEER (Surveillance, Epidemiology and End Results) Program where access to individual patient records is either not possible, or not feasible due to the large number of patients being studied. One alternative would be to adjust the algorithm so that specificity is increased at the expense of sensitivity but this would likely result in a greater number of missed patients. These results are consistent with the notion that the value of text-search methods seems to be in higher sensitivity with lower specificity than other methods of dataset creation using administrative data.⁵¹

A second limitation is the fact that our search was limited to pathology reports at a single institution. This 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 Ottawa Hospital Data Warehouse (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 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. In order 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 a cancer that lies within 15cm of the anal verge.¹⁴⁻¹⁵ 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.

4.2.3 Conclusions and Future Considerations

In conclusion, the present study created and validated a cohort of all rectal cancer resections performed at The Ottawa Hospital 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 sensitivity and specificity of 100% and 98.4%, respectively, resulting in a positive predictive value of 18.6%, meaning each report positively identified by the algorithm had a 18.6% probability of having a diagnosis of rectal cancer. These results demonstrate the feasibility of text-search methods to reliably identify a patient cohort through administrative datasets with high sensitivity. However, in the presence of a low pre-test probability, text-search methods must be combined with a validation method, such as manual chart review, in order to be a viable approach. 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.

4.3 Objective 2: Surgical Quality Indicators

A clear link exists between the quality of surgical resection for rectal cancer and long-term patient outcomes such as local recurrence and survival.²³⁻²⁷ The present study therefore reported the temporal trends of four key surgical quality indicators for all rectal cancer resections performed at The Ottawa Hospital

(TOH) over a 11-year period. The quality audit in the present study yielded several key findings:

1. The number of rectal cancer resections performed at TOH increased over the study period.
2. The success rate for adequate lymph node retrieval (i.e. the proportion of resections where at least 12 lymph nodes were harvested) increased significantly over the course of the study period in both the univariate and multivariate analyses.
3. The proportion of resections with a positive CRM ranged from 2.9-22.2% with an overall mean 6.8% and did not significantly change over the course of the study period.
4. There was no change in the proportion of resections with a positive distal resection margin (DRM), or in the proportion of abdominal perineal resections (APRs) over time.

The clinical relevance in the context of the current literature for each key finding is discussed below.

4.3.1 Increased Rectal Cancer Resections

The number of rectal cancer resections performed at TOH increased significantly from 33 cases in 2000 to 105 cases in 2010 (Figure 9). Several factors likely contributed to this increase, which comes with important clinical implications. First, a steady increase in rectal cancer diagnoses has been demonstrated between the years 2000 and 2015 in other populations and has been attributed to both population growth and an aging population.⁵³ Second, TOH has employed a greater number of colorectal surgeons and surgical oncologists specializing in colorectal cancer resections since the beginning of the study period. In 2000, three colorectal surgeons were employed at TOH with the majority of rectal cancer resections in the region performed by general surgeons without subspecialty training in colorectal surgery. In 2010, 5 surgeons who specialized in the treatment of colorectal surgery were performing the majority of rectal cancer resections. The combination of an increased number of diagnosed rectal cancers along with more colorectal surgery at TOH likely both contributed to the increase in the number of resections over the course of the study period.

Perhaps the most substantial contributor to the rising number of resections was the implementation of a Communities of Practice Surgical Oncology Program (CoP-SOP) at TOH between 2003 and 2005.⁵⁴⁻⁵⁵ The purpose of the CoP-SOP model was to improve regional adherence to accepted treatment algorithms for four types of cancers: Colorectal, Breast, Prostate and Lung.⁵⁵ One of the key components to the model was the formation of a regional cancer centre at TOH that would service the entire Champlain Local Health Integration

Network (LHIN). As of 2004, this program promoted the referral all rectal cancer for the region to this Cancer Assessment Centre (CAC), resulting in a greater proportion of rectal cancer surgery in the region being performed at TOH.⁵⁵ These results show that this objective is being met in the Champlain LHIN.

The overall increase in the number of rectal cancer resections, combined with the implementation of the CoP-SOP model within the Champlain LHIN, has several important clinical implications. First, it triages rectal cancer treatment in the region towards high-volume, subspecialized surgeons who have been shown to produce superior oncologic outcomes in the treatment of rectal cancer.³¹⁻³⁴ Second, the CoP-SOP model facilitates a multidisciplinary cancer conference (MDCC) for each rectal cancer case, which includes input from surgeons, radiologists, oncologists and pathologists.⁵⁵ MDCC's have been shown to improve the accuracy of preoperative staging of rectal cancer as well as adherence to standard treatment guidelines for the disease.⁵⁶ Such conferences may not be possible in smaller, community-based practices.

To date, evidence that the CoP-SOP model in the Champlain LHIN has improved adherence to rectal cancer treatment standards within the region lies in the proportion of patients who receive a preoperative MRI and those referred to radiation and medical oncology.⁵⁷⁻⁵⁸ Therefore, the increased number of rectal cancer resections performed at TOH implies that a greater proportion of patients with rectal cancer are being treated by high-volume surgeons with a multidisciplinary approach. Future research should focus on measuring quality metrics for all rectal cancer patients within the Champlain LHIN.

4.3.2 Lymph Node Retrieval

The present study demonstrated that the proportion of rectal cancer resections at The Ottawa Hospital in which at least 12 lymph nodes were successfully retrieved increased significantly over the course of the study period in both univariate and multivariate analyses. These results are consistent with several institutional and population-based studies that show significant increase in the proportion of colon and rectal surgery cases with successful lymph node yield in both Canada and the United States.⁵⁹⁻⁶¹

Lymph node retrieval has become an important quality indicator for colorectal surgery. Several population-based studies have demonstrated a positive association between the number of lymph nodes harvested and improved survival for both colon and rectal cancer.⁶²⁻⁶⁵ As a result, a minimum harvest of 12 lymph nodes has been endorsed by nearly all of the governing bodies that provide guidelines for the treatment of colorectal cancer including Cancer Care Ontario (CCO), the National Comprehensive Cancer Network (NCCN), and the American Society of Clinical Oncologists (ASCO).^{42,66-67}

The fact that lymph node retrieval is associated with survival⁶²⁻⁶⁵, combined with its ease of measurement and availability on all pathology reports and most colorectal cancer databases, makes it an attractive quality metric for policy makers. In fact, there are several examples in the United States that propose lymph node harvest as a quality measure in certain pay-for-performance models that institute financial penalties for under-performance.⁶⁸ This increased attention to lymph node harvest and clear performance guidelines provided by governing

bodies has undoubtedly contributed to the improvements seen across North America. This is underscored by examples of interventions designed to improve lymph node yield at an institutional level at hospitals across Ontario, including TOH.³⁶

Two key questions arise from this phenomenon. First, if lymph node yield is going to be used as a quality metric, who should be held accountable for underperformance? Second, what is the clinical significance of the improved lymph node yields across North America? To date, these questions remain largely unanswered. Initially it was thought that an inadequate lymph node yield was an indictment on the quality of the surgical resection. However, several regression analyses have identified many independent predictors of lymph node yield in colorectal cancer surgery including factors related to the surgeon, the pathologist, the use of chemo and radiotherapy, and the tumor itself, making the question as to who, or what, is being evaluated unclear.⁶⁹⁻⁷²

Answers to the second question are less clear. As mentioned, there is an association between cancer survival and the number of lymph nodes harvested for both colon and rectal cancer.⁶²⁻⁶⁵ The presumed mechanism for this association was that an inadequate lymph node harvest was more likely to leave cancerous nodes behind, thereby leading to under staging of the disease and depriving the patient of potentially beneficial adjuvant treatments.⁷³ However, improvements in lymph node harvesting have not led to corresponding improvements in patient outcomes. In a national retrospective study, Parsons et al. demonstrated significant improvements in lymph node harvests for colon and

rectal cancer surgery, but no corresponding increase in the diagnosis of higher stage tumors and no improvements in overall patient survival.⁵⁹ This has brought into question the under staging mechanism theory, providing evidence that perhaps a confounding variable associated with both lymph node yield and survival is the true cause. One such theory is that lymph node yield is a marker that reflects the patient's immune response to the tumor, with higher lymph node counts representing a more pronounced tumor-host response, thereby leading to better outcomes.⁷²

In conclusion, we demonstrated significant improvement in the proportion of successful lymph node retrieval for rectal cancer surgery at TOH over the 11-year study period in both univariate and multivariate analyses. This is consistent with similar findings across North America and has likely resulted from increased attention paid to lymph node harvest as a surgical quality metric. However, the utility of using lymph nodes as measure of surgical quality remains unclear. Further research should focus on whether or not improving lymph node yield will actually result in improved patient-care as opposed to being used as a quality metric because it is readily available information and easy to measure.

4.3.3 Circumferential Resection Margin

Our results demonstrated the proportion of resections with a positive CRM ranged from 2.9-22.2% with an overall mean 6.8% and did not significantly change over the course of the study period. These results are consistent with, if not better than, reported CRM positivity rates in Canada and internationally.

DeCaria et al. reported provincial CRM positivity rates in Canada ranged from 7.7% to 21.1% between the years 2009 and 2010.⁷⁴ Nagtegaal and Quirke reported similar rates in the United Kingdom.⁷⁵

Circumferential resection margin is a valuable quality indicator in rectal cancer surgery. Since the early 1980's it has consistently been shown to be an independent predictor of local recurrence and survival in patients undergoing a rectal cancer resection.^{18-19,76} CRM positivity has also been shown to vary between surgeons, making it a useful indicator for the quality of the surgical resection.⁷⁶

Even though the overall mean CRM positivity rate in our study is in line with published reports from other institutions and populations, several other explanations exist for the lack of improvement over time. First, the use of preoperative MRI in rectal cancer patients and neoadjuvant chemoradiation for higher stage tumors have both been shown to impact CRM positivity rates.⁷⁷⁻⁷⁸ Second, several studies have demonstrated improved outcomes, including lower CRM positivity rates, for rectal cancer patients who have their surgery performed at a high-volume institution with high volume surgeons.³¹⁻³⁴ Since the implementation of the Communities of Practice Surgical Oncology Program (CoP-SOP) model at TOH in 2003 adherence to guidelines for preoperative imaging and appropriate use neoadjuvant chemoradiation have been excellent.⁵⁸ Similarly, all of the rectal cancer resections were performed by high-volume surgeons with subspecialty training in the treatment of colorectal cancer.⁵⁶⁻⁵⁷ Therefore, it is possible that CRM positivity rates for resections performed at

TOH have remained stable because of ongoing high quality care that adheres to accepted guidelines.

Finally, several tumor and patient factors that independently predict positive CRM have been identified and were adjusted for in the present study. These include higher stage tumors, low tumors close to the anal verge, multivisceral resections (MVRs) and abdominal perineal resections (APRs), and resections performed on an emergency basis.⁷⁹ The present study adjusted for all of these variables using a multivariable logistic regression analysis, which may help to explain the lack of change in CRM positivity over the course of the study period.

In conclusion, our results demonstrated excellent overall CRM positivity rates for rectal cancer resections with no significant change over the course the entire of the study period. This is likely explained by the centralization of rectal cancer surgery to high volume surgeons and the multidisciplinary approach supported by the CoP-SOP framework implemented in 2003. Future research could focus on CRM positivity on a population-scale as efforts similar to CoP-SOP model look to improve surgical quality for rectal cancer across the region and province.

4.3.4 Distal Resection Margin and APR Rate

The final two quality indicators that were measured were the proportion of cases with positive distal resection margin (DRM) and the proportion of cases performed as an abdominal perineal resection (APR). The proportion of cases with a positive DRM were very low, with 7 out of 11 years having 0 positive

DRM's, with the highest rate occurring in 2008 at 2.6%. Multivariate logistic regression analysis was not performed due to the small number of events over the course of the study period. Similarly, the proportion of cases performed as APR's did not significantly change over the over the course of the study period, but the overall proportion was much higher, ranging from 26.1%-41.9% with an overall mean of 33.8%

Positive DRM has the same implication as positive CRM with respect to local recurrence and survival.²³⁻²⁴ However, the rate of positive DRM tends to be much lower. Positive margins in any rectal cancer operation are directly associated with its technical challenges. The circumferential margin can often be abutting, or invading, important surrounding structures within the pelvis including the pelvic sidewall, the prostate and seminal vesicles in men, and the vagina in women. Resecting a locally advanced tumor with negative margins while protecting these surrounding structures can be technically challenging and explains the higher rates of positive CRM compared to DRM.

Little has been published on the actual rates of positive DRM for rectal cancer in Ontario or other jurisdictions. This is likely due to the changing nature of what constitutes DRM over the years. Initially, a distance 5cm was recommended for both proximal and distal resection margins for colon and rectal cancer.⁸⁰ However, this may not be possible for rectal tumors located in the distal third of the rectum, thereby necessitating an APR. Given the benefits of sphincter sparing operations, several studies have investigated the need for such a wide margin. To date no oncologic benefit to a distal resection margin greater than

2mm has been demonstrated.⁸¹ Currently, Cancer Care Ontario recommends an ideal distal resection of 2cm, but states that “a negative margin of less than 2cm can be oncologically adequate to facilitate very low colorectal re-anastomosis.”⁴²

Although the actual rates of positive DRM on a population scale are not generally reported, what is clear is that a negative distal resection margin should not be compromised in an effort to avoid an APR.⁴² Therefore, positive DRM rates should be exceedingly low. We feel the overall proportion of positive DRM of 0.5% reported in the present study is consistent with this principle.

APR's have been identified as an independent predictor of increased local recurrence and decreased disease-free survival when compared to anterior resections.²⁷ This is likely a direct result of the fact that APR's have been independently associated with higher rates of positive CRM when compared to low anterior resections (LAR) while controlling for important covariates such as tumor stage, tumor location, and use of neoadjuvant chemoradiation.⁸² Therefore, limiting the proportion of APR's performed for all rectal cancer resections is an important quality measure.

Several factors are associated with the decision to perform an APR including the distance from the anal verge, the tumor depth (T-stage), and involvement of surrounding pelvic structures. Therefore, it is important to control for these factors when looking at APR's as a quality measure. Our results compare favourably with published reports while controlling for these factors.⁸³

A revised technique for the abdominal perineal resection has recently been proposed.⁸⁴ Traditionally the mesorectum is dissected off of the levator

muscles during the perineal portion of the operation. The newer, extended APR promotes an en bloc resection of the entire levator muscle complex, using a myocutaneous flap to reconstruct the perineal defect.⁸⁴ This is referred to as an extra-levator abdominal perineal excision (ELAPE). Several retrospective studies have suggested ELAPE is associated with reduced CRM involvement and local recurrence rates when compared to traditional APR.⁸⁵⁻⁸⁶ One prospective study has been performed, and while it demonstrated acceptable positive CRM rates for ELAPE, it did not directly compare to the traditional APR technique.⁸⁷ Therefore, until the APR has been reliably shown to have similar oncologic outcomes when compared to low anterior resections, it should be used as an important quality indicator for the treatment of rectal cancer in addition to CRM positivity.

Conclusions: In conclusion, both DRM positivity and the proportion of APR's are important surgical quality indicators for rectal cancer treatment. The rates for both in the present study compared well with published results. APR should continue to be measured as a quality measure in the treatment of rectal cancer until an alternative such as ELAPE is clearly established to have similar results to the low anterior resection.

4.3.5 Strengths and Limitations

There are several strengths and limitations to the surgical quality audit performed in Objective 2 of the present study. The first, and most important, strength is the audit itself. We believe that the validated cohort developed in Objective #1 of the study contains all patients with a rectal cancer resection performed at TOH over the course of the study period and the benefits to auditing surgical quality indicators for rectal cancer have been clearly demonstrated. Several population-based audits of a range of rectal cancer quality indicators and outcomes performed in several European countries resulted in significant improvement in many outcomes including 30-day mortality rates, proportion of successful TME excision, local recurrence rates and long-term mortality.⁸⁸⁻⁹⁰ In addition, the ability to audit all patients allows for the inclusion of marginalized portions of a population that often get excluded from prospective studies and randomized trials, such as the elderly and those with significant comorbidities.⁹¹

A second strength of the study is the quality of the data. All data was extracted directly from patient charts. In particular, each of the four surgical quality measures studies was directly available within the text of the pathology report. The result was no missing data for any of the surgical quality measures. The data quality for the covariates was nearly as good, as it was also extracted directly from patient charts, including clinic notes, operative reports, and radiology reports. Traditionally, large scale audits of populations using administrative datasets compile patient information and outcomes using

diagnostic and procedure codes, the accuracy of which can be questioned.³⁸ By extracting data through manual chart review, we can be confident in not only the completeness of the data, but also the accuracy.

Several study limitations also exist. First, we performed a surgical quality audit of only a single, tertiary care centre. This not only limits the generalizability of the results obtained, it also may cause misinterpretation. For example, one of the main tenets of the CoP-SOP model initiated in the Champlain LHIN in 2003 was the centralized referral system for all rectal cancer patients to the tertiary care centre at TOH.⁵⁵ This resulted in a greater proportion of rectal cancer surgery in the LHIN being performed by high-volume surgeons using a multidisciplinary approach, both of which have been shown to be associated with improved outcomes.^{31-34,56} By only tracking quality measures at TOH, we may be missing substantial quality improvements in the LHIN over the course of the study period. That is, quality at TOH may not improve because it was already high, and by centralizing the referral system, a greater proportion of patients in the region may be receiving better care. Such a finding would be missed in our study.

Another limitation is the somewhat narrow scope of the quality indicators that were chosen. The four surgical quality measures in our study, while valid, only address direct outcomes of the surgeries themselves, such as margins and the number of lymph nodes retrieved. Several other, non-surgical quality measures have also been identified as important in the care of rectal cancer such as adequate pre-operative imaging, appropriate referral to radiation and medical

oncology, and discussion of cases at multidisciplinary cancer conferences.⁹²

Future research should look to combine both operative and non-operative quality care measures for the treatment of rectal cancer.

Finally, while the present study adjusted the majority of important covariates when assessing quality measures, tumor circumferential location was not included. The location of the tumor within the lumen of the rectum can affect patient outcomes. For example, an anterior location has been independently associated with increased rates of positive CRM and local recurrence.⁹³ This is likely due to the short distance between the anterior mesorectum and important surrounding structures such as the seminal vesicles and prostate in men and the vagina in women. While tumor distance from the anal verge was readily available in patient records, circumferential location was not.

4.3.6 Conclusions and Future Considerations

In conclusion, we performed an audit of four key pathological outcomes for the treatment of rectal cancer. We demonstrated significant increase in the number of rectal cancer resections performed at TOH as well as the proportion of cases with successful lymph node retrieval. Performance for the remaining three surgical quality measures was acceptable when compared to previously published population-based results, and did not show significant change over the course of the study period. Future research for the quality of rectal cancer care at TOH and in the Champlain LHIN should include results from all hospitals within the region to assess the overall impact of the CoP-SOP program. Similarly, non-

surgical quality measures should also be combined with the surgical measures described in this study.

4.4 Objective 3: Validity Test of Procedure Codes Identifying Rectal Cancer

4.4.1 Discussion of Findings in Context of Current Literature

The present study assessed the validity of a previously published algorithm to identify patients who underwent a surgical resection for rectal cancer using population-based, administrative datasets. When compared to the gold-standard cohort of patients created in Objective #1 of this study, this algorithm successfully identified rectal cancer patients with a sensitivity and specificity of 89.5% and 99.9%, respectively. With a disease prevalence of 0.12% in the study population, this resulted in a positive predictive value (PPV) of 64.9%. This means that greater than 1 in 3 patients identified by this algorithm were incorrectly labeled as having a rectal cancer resection.

The accuracy of coding algorithms to identify patient cohorts through administrative data varies. Widdifield et al. tested over 100 coding algorithms used to identify patients with rheumatoid arthritis and found that the PPV ranged between 51% and 83%, which was thought to represent adequate accuracy.⁹⁴ Another study that used procedure codes to identify patients from the CIHI discharge abstract database (DAD) in Canada demonstrated PPV's of 95%, 96%, and 98% for percutaneous coronary intervention (PCI), cardiac catheterization, and coronary artery bypass graft (CABG) surgery, respectively.⁹⁵

Little has been reported on the accuracy of administrative datasets for identifying patients who have undergone rectal cancer surgery, specifically. Goldsbury et al. reported the accuracy for identifying rectal cancer surgery within a population-based cancer registry in New South Wales, Australia. They demonstrated a sensitivity and specificity of 85% and 96%, respectively, which are similar to our results.⁹⁶ However, this study validated that data within a prospectively collected cancer registry whose sole purpose is to record demographic, treatment and outcomes data for cancer patients within a population. Such a registry is distinctly different from administrative data, which is defined as data collected and abstracted for administrative, and not clinical, purposes.³⁷

Li et al. reported results that are directly comparable to the present study.⁹⁷ In this study, the authors compared the accuracy of a coding algorithm utilizing hospital inpatient population-based datasets in the province of Alberta, Canada (datasets that include ones similar to the CIHI-DAD, as well as physician billing data) to a gold-standard patient cohort for identifying patients who had a surgical resection for rectal cancer. The gold-standard population in this study was from a prospectively collected cancer registry for all patients in the province of Alberta, Canada, which has been thoroughly validated and commended for the quality of its data.⁹⁸ The authors demonstrated a sensitivity and specificity of 97% and 60%, respectively. While these results are not dissimilar to ours, the PPV reported was much higher at 93%, compared to 64.9% in our study.

These results are consistent with those reported in the present study. The sensitivity and specificity of 89.5% and 99.9%, respectively, compare very favourably. Although the PPV is markedly lower in our study, it is important to note that this is dependent on disease prevalence. In our study, the denominator used for validating the algorithm was the total number of admissions at TOH over the study period present within ICES datasets, which was 664,075. This resulted in a disease prevalence of rectal cancer resections at TOH over the course of the study period of 0.12%. Li et al first identified their patient cohort through the Alberta Cancer Registry, which would have a much higher pre-test probability for rectal cancer surgery than other, non-cancer based datasets, thereby potentially explaining the higher PPV when compared to our study.⁹⁷

It is also possible, however, that the coding algorithm validated in our study is simply less accurate at identifying rectal cancer patients. One key difference between the two studies is the use of physician billing codes. The algorithm validated in our study does not incorporate billing codes while the Li study does.⁹⁷ A more comprehensive coding algorithm could therefore explain the difference in PPV between the two studies.

4.4.2 Strengths and Limitations

This study has several strengths, the first of which is the mere practice of validating a coding algorithm used in administrative data research. Very few studies that use administrative data utilize validated coding algorithms to develop their patient cohorts. In a study that evaluated 115 randomly selected, peer-

reviewed, published studies that used codes to define populations, key variables, or outcomes, only 14 (12.2%) cited any statistic measuring the code's association to the entity it represented.³⁸ Only 5 (4.3%) studies provided enough information where this association could be measured, and in 2 of the 5 this association was less than 50%.³⁸

A second strength is the fact that this study validates the complete methodology used to create the patient cohort, rather than just the codes themselves. For example, the CCI codes used in ICES databases to identify colorectal surgical resections have been validated against chart extraction, and were found to be very reliable with a sensitivity of 94% and a PPV of 90% (ICES Report).⁹⁹ However, simply using validated procedure or diagnostic codes does not necessarily ensure that an entire algorithm used to extract data from large, administrative datasets is also valid. The present study examines the actual accuracy of such an algorithm, providing a probability that a patient identified from that method actually has the disease in question. This has a much more practical application for researchers who use administrative data, providing a reliable method for identifying a specific patient cohort.

Our study also has several weaknesses. First and foremost is that we only used a single institution to validate the previously described coding algorithm. The algorithm tested attempted to identify all rectal cancer resections in the province of Ontario using administrative datasets available through ICES.⁴⁰ While the present study validated the accuracy of this algorithm, we did so using a gold standard cohort from only one institution within the province. This can limit the

generalizability of the results. Other institutions that perform rectal cancer surgery in Ontario may have more or less reliable data entry methods into administrative datasets such as CIHI. A more complete validation study would therefore utilize at least a sample of rectal cancer surgery cases from multiple institutions around the province, thereby making the results more generalizable.

A second limitation is the assumption that the gold-standard cohort created in Objective #1 is, in fact, 100% accurate. Since each chart in this cohort was manually reviewed, we can be reasonably assured that there are no false-positives within this validation cohort. However, we assumed 100% sensitivity based on testing performed in Objective #1 on the text-search algorithm used to capture rectal cancer pathology reports (Tables 2 & 3). We assumed that this 100% sensitivity carried over when applied to all pathology reports in the OHDW. If this was incorrect, it could have resulted in some missed pathology reports, compromising the accuracy of the gold standard, validation cohort, thereby skewing the results of the ensuing validation study performed in Objective #3.

Finally, the present study provides a measure of the accuracy of previously described coding algorithms for identifying rectal cancer resections through ICES databases. Specifically, we calculated that the previously described algorithm had a sensitivity and specificity of 89.5% and 99.9%, respectively, as well as a positive predictive value of 64.9%. While these values represent adequate accuracy when compared to similar validation studies, we do not provide any methods for improving upon the tested algorithm. Van Walraven et al. demonstrated the ability to improve the accuracy of methods used to

identify patient cohorts in administrative datasets using a logistic regression model that combines multiple variables available within datasets.¹⁰⁰ Such a model could be used to improve upon the 64.9% PPV demonstrated in the present study. This could potentially provide a more useful, validated method for accurately identifying patients with rectal cancer surgery on a population scale in the province of Ontario using ICES databases.

4.4.3 Conclusions Future Considerations

In conclusion, Objective #3 of the present study tested the validity of a previously published method for identifying rectal cancer surgery patients in the province of Ontario using administrative databases available in ICES. When tested against the gold standard, validation cohort created in Objective #1 of the present study, these methods were found to have a sensitivity and specificity of 89.5% and 99.9% respectively, with a PPV of 64.9% and a NPV of 99.9%. This suggests that the probability that patients identified using this previously published algorithm actually had a surgical resection for rectal cancer was 64.9%. While these results are not dissimilar to values of similar studies, they indicate that more than a third of rectal cancer resections identified using procedure codes in administrative datasets are incorrectly identified. Future research should attempt to improve on these results utilizing other variables available within administrative datasets. The result could be the development of a highly accurate, validated method for identifying rectal cancer resections on a population scale

within the province of Ontario, which could have significant implications for researchers who wish to study this group.

References

1. The Canadian Cancer Society. Colorectal cancer statistics.
<http://www.cancer.ca/en/cancer-information/cancer-type/colorectal/statistics/?region=qc>. Updated 2013. Accessed 07/14, 2013
2. Moore KL, Agur AMR. Pelvis and perineum. In: Sun B, ed. *Essential clinical anatomy*. 2nd ed. Baltimore, MD: Lippincott Williams & Wilkins; 2002:247-252.
3. Godwin JD and Brown C. Some prognostic factors in survival of patients with cancer of the colon and rectum. *J Chron Dis*. 1975; 28:441-454
4. Wolmark N, Wieand HS, Rockette HE 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(6): 743-752.
5. O'Connell JB, Maggard MA, Ko CY. Colon cancer survival rates with the new American Joint Committee on cancer sixth edition staging. *Journal of the National Cancer Institute*. 2004; 96(19): 1420-1425.
6. AJCC (American Joint Committee on Cancer) Cancer Staging Manual, 7th edition, Edge, SB, Byrd, DR, Compton, CC, et al (Eds) (Eds), Springer, New York 2010.
p.143
7. Wong R, Berry S, Spithoff K et al. Preoperative or Postoperative Therapy for the Management of Patients with Stage II or III rectal cancer: Guideline Recommendations. July, 2008. Available at
<https://www.cancercare.on.ca/cms/One.aspx?portalId=1377&pageId=10207>.
Accessed 08/07, 2013.

8. Tjandra JJ, Kilkenny JW, Buie WD et al. Practice parameters for the management of rectal cancer (revised). *Dis Colon Rectum*. 2005; 48:411-423.
9. Jonker D, Spithoff K, Maroun J and the Gastrointestinal Cancer Disease Site Group. Adjuvant systemic chemotherapy for stage II and III colon cancer following complete resection. April, 2008. Available at <https://www.cancercare.on.ca/cms/One.aspx?portalId=1377&pageId=10207>. Accessed 08/07, 2013.
10. Milson JW, Bohm B, Hammerhofer KA, Fazio V, Steiger E, Elson P. A prospective, randomized trial comparing laparoscopic versus conventional techniques in colorectal cancer surgery: a preliminary report. *J Am Coll Surg*. 1998; 187(1): 46-54.
11. Guillou PJ, Quirke P, Thorpe H et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patients with colorectal cancer (MRC CLASICC trial): multicenter, randomised controlled trial. *Lancet*. 2005;365:1718-1726.
12. Lujan J, Valero G, Hernandez Q, Sanchez A, Frutos MD, Parilla P. Randomized clinical trial comparing laparoscopic and open surgery in patients with rectal cancer. *British Journal of Surgery*. 2009;96:982-989.
13. Araujo SEA, da Silva AH, de Campos FG et al. Conventional approach x laparoscopic abdominoperineal resection for rectal cancer treatment after neoadjuvant chemoradiation: results of a prospective randomized trial. *Rev Hosp Clin Fac Med*. 2003; 58(3):133-140.

14. Morris EJ, Sandin F, Lambert PC et al. A population-based comparison of the survival of patients with colorectal cancer in England, Norway and Sweden between 1996 and 2004. *Gut*. 2011; 60:1087-1093.
15. Ciccolallo L, Capocaccia R, Coleman MP et al. Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut*. 2005; 54:268-273.
16. Halvorsen TB, Johannesen E. DNA ploidy, tumor site, and prognosis in colorectal cancer: a flow cytometric study of paraffin-embedded tissue. *Scand J Gastroenterol*. 1990; 25:141-148.
17. Grinnell RS. The spread of carcinoma of the colon and rectum. *Cancer*. 1950; 3(4): 641-652.
18. Quirke P, Dixon MF, Durdey P, Williams NS. Local recurrence of rectal adenocarcinoma due to inadequate surgical resection: histopathological study of lateral tumor spread and surgical excision. *Lancet* 1986; 328:996-9.
19. Quirke P, Dixon MF. The prediction of local recurrence in rectal adenocarcinoma by histopathological examination. *Int J Colorectal Dis* 1988; 3:127-31.
20. Heald RJ, Husband EM, Ryall EDH. The mesorectum in rectal cancer surgery – the clue to pelvic recurrence? *British Journal of Surgery*. 1982; 69:613-616.
21. Simunovic M, Goldsmith C, Thabane L, McLeod R, Denardi F, Whelan TJ, Levine MN. The Quality Initiative in Rectal Cancer (QIRC) trial: study protocol of a cluster randomized controlled trial in surgery. *BMC Surgery* 2008; 8:4-11.

22. Kapiteijn E, Putter H, van de Velde CJH. Impact of the introduction and training of total mesorectal excision on recurrence and survival in rectal cancer in The Netherlands. *British Journal of Surgery* 2002; 89:1142-1149.
23. Kim YW, Kim NK, Min BS et al. Factors associated with anastomotic recurrence after total mesorectal excision in rectal cancer patients. *J Surg Oncol.* 2009; 99(1):58-64.
24. Leo E, Belli F, Miceli R et al. Distal clearance margin of 1 cm or less: a safe distance in lower rectum cancer surgery. *Int J Colorectal Dis.* 2009; 24(3): 317-322.
25. Berberoglu U. Prognostic significance of total lymph node number in patients with T1-4N0M0 colorectal cancer. *Hepatogastroenterology.* 2004; 51(60): 1689-1693.
26. Swanson RS, Compton CC, Stewart AK, Bland KI. The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined. *Ann Surg Oncol.* 2003; 10(1):65-71.
27. den Dulk M, Collette L, van de Velde CJ et al. Quality of surgery in T3-4 rectal cancer: involvement of circumferential resection margin not influenced by preoperative treatment. Results from EORTC trial 22921. *Eur J Cancer.* 2007; 43(12):1821-1828.
28. Sant M, Allemani C, Santaquilani M, Knijn A, Marchesi F, Ca-pocaccia R. EURO-CARE-4. Survival of cancer patients diagnosed in 1995-1999. Results and commentary. *Eur J Cancer.* 2009;45(6):931-91.

29. Ciccolallo L, Capocaccia R, Coleman MP, Berrino F, Coebergh JWW, Damhuis RAM, Faivre J, Martinez-Garcia C, Moller H, Ponz de Leon M, Launoy G, Raverdy N, Williams EMI, Gatta G. Survival differences between European and US patients with colorectal cancer: role of stage at diagnosis and surgery. *Gut* 2005; 54:268-273.
30. Bjerkeset T, Edna T-H. Rectal cancer: the influence of type of operation on local recurrence and survival. *Eur J Surg* 1996; 162:643-8.
31. Porter GA, Soskoine CL, Yakimets WW, Newman SC. Surgeon-related factors and outcome in rectal cancer. *Annals of Surgery* 1998; 227(2):157-167.
32. Drolet S, MacLean R, Myers P, Shaheen AM, Dixon E, Buie WD. Elective resection of colon cancer by high volume surgeons is associated with decreased morbidity and mortality. *J Gastrointest Surg*. 2001; 15:541-550.
33. Karanicolos PJ, Dubois L, Colquhoun PHD, Swallow CJ, Walter SD, Guyatt GH. The more the better? The impact of surgeon and hospital volume on in-hospital mortality following colorectal resection. *Ann Surg*. 2009; 249(6):954-959.
34. Hannan EL, Radzyner MR, Rubin D, Daugherty J, Brennan MF. The influence of hospital and surgeon volume on in-hospital mortality for colectomy, gastrectomy, and lung lobectomy in patients with cancer. *Surgery*. 2002; 131(1):6-15.
35. Fung-Kee-Fung M, Goubanova E, Sequeira K, Abdulla A et al. Development of communities of practice to facilitate quality improvement initiatives in surgical oncology. *Q Manage Health Care* 2008; 17(2):174-185.

36. Smith AJ, Law CHL, Khalifa MA et al. Multimodal CME for surgeons and pathologists improves colon cancer staging. *Journal of Cancer Education*. 2003; 18(2):81-85.
37. The Center for Administrative Data Research. <http://cadr.wustl.edu>. Updated 2013. Accessed on 08/08, 2013.
38. van Walraven C, Bennett C, Forster AJ. Administrative database research infrequently uses validated diagnostic or procedural codes. *J Clin Epidemiol*. 2011; 64:1054-1059.
39. Chan BP, Gomes T, Musselman RP, Auer RC, Moloo H, Poulin EC, Mamazza J, Al-Khayal K, Mamdani M, Al-Omran M, Boushey RP, Al-Obaid O. Trends in Colon Cancer Surgery in Ontario: 2002-2009. *Colorectal Disease*. 2012; 14(10):e708-712.
40. Musselman RP, Gomes T, Chan BP, Auer RC, Moloo H, Poulin EC, Mamazza J, Al-Khayal K, Mamdani M, Al-Omran M, Al-Obaid O, Boushey RP. Changing Trends in Rectal Cancer Surgery in Ontario: 2002-2009. *Colorectal Disease*. 2012; 14(12):1467-1472.
41. Faiz O, Warusavitarne J, Bottle A et al. Laparoscopically assisted vs. open elective colonic and rectal resection: a comparison of outcomes in English National Health Service Trusts between 1996 and 2006. *Dis Colon Rectum* 2009; 52:1695-1704.
42. Smith AJ, Driman DK, Spithoff K, McLeod R, Hunter A, Rumble RB, Langer, B. Optimization of surgical and pathological quality performance in radical surgery for colon and rectal cancer: margins and lymph nodes. Available at: [http://www.cancercare.on.ca/toolbox/qualityguidelines/clin-program/surgery-abs/](http://www.cancercare.on.ca/toolbox/qualityguidelines/clin-program/surgery-<u>abs</u>/). Updated 09/2008. Accessed 08/17, 2013.

43. Van der Pas MH, Haglind E, Cuesta MA et al. Laparoscopic versus open surgery for rectal cancer (COLOR II): short-term outcomes of a randomised, phase 3 trial. *Lancet Oncol.* 2013; 14(3):210-218.
44. Guillou PJ, Quirke P, Thorpe H et al. Short-term endpoints of conventional versus laparoscopic-assisted surgery in patient with colorectal cancer (MRC CLASICC trial): multicentre, randomised controlled trial. *Lancet.* 2005;365:1718-1726.
45. Virnig BA, McBean M. Administrative data for public health surveillance and planning. *Annu Rev Public Health* 2001; 22:213-30.
46. Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. *J Public Health Med* 2001;23:205-211.
47. Peabody JW, Luck J, Jain S, Bertenthal D, Glassman P. Assessing the accuracy of administrative data in health information systems. *Med Care* 2004;42:1066-1072.
48. Hohl CM, Karpov AK, Reddekopp L, Stausberg J. ICD-10 codes used to identify adverse drug events in administrative data: a systematic review. *J Am Med Inform Assoc* 2014; 21:547-557.
49. Nelson HD, Weerasinghe R, Bifulco C, Assur T, Elmore JG, Weaver DL. Development of an electronic breast pathology database in a community health system. *J Pathol Inform* 2014; 30:5-26.
50. Ivers N, Pylypenko B, Tu K. Identifying patients with ischemic heart disease in an electronic medical record. *J Prim Care Community Health* 2011; 2(1):49-53.
51. Harvey JM, Fitzhenry F, Matheny ME et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011; 306(8):848-855.

52. McKenzie K, Scott DA, Campbell MA, McClure RJ. The use of narrative text for injury surveillance research: a systematic review. *Accid Anal Prev* 2010; 42:354-363.
53. Cancer Care Ontario. Ontario cancer plan: 2005 progress report. Toronto, Ontario, Canada: *Cancer Care Ontario*; 2005.
54. Gagliardi A, Ashbury F, George R, Irish J, Stern H. Improving cancer surgery in Ontario: recommendations from a strategic planning retreat. *Can J Surg* 2004; 47(4):270-276.
55. Stern H. Cancer surgery standards: a regional quality improvement strategy. Available at <https://www.cancercare.on.ca/common/pages/UserFile.aspx?fileId=13752>. Accessed 08/01, 2015.
56. Snelgrove RC, Subendran J, Jhaveri K et al. Effect of multidisciplinary cancer conference on treatment plan for patients with primary rectal cancer. *Dis Colon Rectum* 2015; 58(7):653-658.
57. Williams LJ, Morash R, Shin S et al. Successful implementation of a novel communities of practice (CoP) model to facilitate quality improvement initiatives in colorectal cancer surgery. *Canadian Surgery Forum Abstracts* 2010; #143.
58. Williams L, Crawford A, McLaughlin K et al. Implementation of a regional colorectal cancer communities of practice (CoP) leads to improved quality of care for rectal cancer patients. *Canadian Surgery Forum Abstracts* 2011; #109.
59. Parsons HM, Tuttle TM, Kuntz KM, Begun JW, McGovern PM, Virnig BA. Association between lymph node evaluation for colon cancer and node positivity over the past 20 years. *JAMA* 2011; 306(10):1089-1097.

60. Johnson PM, Malatjalian D, Porter GA. Adequacy of nodal harvest in colorectal cancer: a consecutive cohort study. *J Gastrointest Surg* 2002; 6:883-890.
61. Gonsalves WI, Kanuri S, Tashi T et al. Clinicopathologic factors associated with lymph node retrieval in resectable colon cancer: a veterans' affairs central cancer registry (VACCR) database analysis. *J Surg Oncol* 2011; 104:667-671.
62. Wong JH, Bowles BJ, Beuno R, Shimizu D. Impact of the number of negative nodes on disease-free survival in colorectal cancer patients. *Dis Colon Rectum* 2002; 45(10):1341-1348.
63. Le Voyer TE, Sigurdson ER, Hanlon AI et al. Colon cancer survival is associated with increasing number of lymph nodes analyzed: a secondary survey of intergroup trial INT-0089. *J Clin Oncol* 2003; 21(15):2912-2919.
64. Swanson RS, Compton CC, Stewart AK, Bland KI. The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined. *Ann Surg Oncol* 2003; 10(1):65-71.
65. Chang GJ, Rodriguez-Bigas MA, Skibber JM, Moyer VA. Lymph node evaluation and survival after curative resection of colon cancer: systematic review. *J Natl Cancer Inst* 2007; 99(6):433-441
66. National Quality Forum: Specifications of the national consensus standards for breast and colon cancer. Available at:
https://www.qualityforum.org/News_And_Resources/Press_Releases/2012/NQF_Endorses_Additional_Cancer_Measures.aspx. Last Accessed 08/01, 2015.
67. Nelson H, Petrelli N, Carlin A et al. Guidelines 2000 for colon and rectal cancer surgery. *J Natl Cancer Inst* 2001; 93(8):582-96.

68. Institute of Medicine of The National Academies. Performance measurement accelerating improvement. Washington (DC): *The National Academies Press*; 2006.
69. Senthil M, Trisal V, Paz IB, Lai LL. Prediction of the adequacy of lymph node retrieval in colon cancer by hospital type. *Arch Surg* 2010; 145(9):840-843.
70. Chao-Wen H, Chieh-Hsin L, Jui-Ho W et al. Factors that influence 12 or more harvested lymph nodes in early-stage colorectal cancer. *World J Surg* 2009; 33:333-339.
71. Jakub JW, Russell G, Tillman CL, Laiscy C. Colon cancer and low lymph node count: who is to blame? *Arch Surg*. 2009; 144(12): 1115-1120.
72. Ricciardi R, Baxter N. Association versus causation versus quality improvement: setting benchmarks for lymph node evaluation in colon cancer. *J Natl Cancer Inst* 2007; 99(6):414-415.
73. Joseph NE, Sigurdson ER, Hanlon AL, Wang H, Mayer RJ, MacDonald JS, et al. Accuracy of determining nodal negativity in colorectal cancer on the basis of the number of nodes retrieved on resection. *Ann Surg Oncol* 2003; 10 : 213 – 8.
74. DeCaria K, Rahal R, Niu J, Lockwood G, Bryant H. Rectal cancer resection and circumferential margin rates in Canada: a population-based study. *Curr Oncol* 2015; 22(1): 60-63.
75. Nagtegaal ID, Quirke P. What is the role for the circumferential margin in the modern treatment of rectal cancer? *J Clin Oncol* 2008; 26:303-312.
76. Birbeck KF, Macklin CP, Tiffin NJ et al. Rates of circumferential resection margin involvement vary between surgeons and predict outcomes in rectal cancer surgery. *Ann Surg* 2002; 235(4):449-457.

77. MERCURY Study Group. Diagnostic accuracy of preoperative magnetic resonance imaging in predicting curative resection of rectal cancer: prospective observational study. *BMJ*, doi:10.1136/bmj.38937.646400.55 (published 25 September 2006)
78. Fokesson J, Birgisson H, Pahlman L, Cedermark B, Glimelius B, Gunnarsson U. Swedish rectal cancer trial: long lasting benefits from radiotherapy on survival and local recurrence rate. *J Clin Oncol* 2005; 23:5644-5650.
79. Hiranyakas A, da Silva G, Wexner SD, Ho YH, Allende D, Berho M. Factors influencing circumferential resection margin in rectal cancer. *Colorectal Dis* 2012; 15(3):298-303.
80. Scott N, Jackson P, Al-Jaberi T, Dixon MF, Quirke P, Finan PJ. Total mesorectal excision and local recurrence: a study of tumour spread in the mesorectum distal to rectal cancer. *Br J Surg*. 1995; 82:1031–1033.
81. Bujko K, Rutkowski A, Chang GJ, Michalski W, Chmielik E, Kusnierz. Is the 1-cm rule of distal bowel resection margin in rectal cancer based on clinical evidence? A systematic review. *Ann Surg Oncol*. 2012; 19(3):801-808.
82. den Dulk M, Putter H, Collette L et al. The abdominoperineal resection itself is associated with an adverse outcome: The European experience based on a pooled analysis of five European randomized clinical trials on rectal cancer. *Eur J Cancer*. 2009; 45:1175-1183.
83. Musselman RP, Gomes T, Chan BP et al. Changing trends in rectal cancer surgery in Ontario: 2002-2009. *Colorectal Dis*. 2012; 14:1467-1472.
84. Holm T, Ljung A, Haggmark T, Jurrell G, Lagergren J. Extended abdominoperineal resection with gluteus maximus flap reconstruction of the pelvic floor for rectal

- cancer. *Brit J Surg*. 2007; 246:83-90.
85. West NP, Anderin C, Smith KJ, Holm T, Quirke P. European extralevator abdominoperineal excision for low rectal cancer. *Br J Surg*. 2010; 97:588-599.
86. Stelzner S, Koehler C, Stelzer J, Sims A, Wtzigimann H. Extended abdominoperineal excision vs. standard abdominoperineal excision in rectal cancer: a systematic overview. *Int J Colorectal Dis*. 2011; 26:1227-1240.
87. Han JG, Wang ZJ, Qian Q et al. A prospective multicenter clinical study of extralevator abdominoperineal resection for locally advanced low rectal cancer. *Dis Colon Rectum*. 2014; 57(12):1333-1340.
88. Birgisson H, Talback M, Gunnarsson U, Pahlman L, Glimelius B. Improved survival in cancer in the colon and rectum in Sweden. *Eur J Surg Oncol*. 2005; 31(8):845-853.
89. Harling H, Bulow S, Kronberg O, Moller LN, Jorgensen T. Survival of rectal cancer patients in Denmark during 1994-99. *Colorectal Dis*. 2004; 6(3):153-157.
90. Wibe A, Moller B, Norstein J, et al. A national strategic change in treatment policy for rectal cancer – implementation of total mesorectal excision as routine treatment in Norway. A national audit. *Dis Colon Rectum*. 2002; 45(7):857-866.
91. van Gijn W, van de Velde CJH. 2010 SSO John Wayne clinical research lecture: rectal cancer outcome improvements in Europe: population-based outcome registrations will conquer the world. *Ann Surg Oncol*. 2011; 18:691-696.
92. Vergara-Fernandez O, Swallow CJ, Victor JC et al. Assessing outcomes following surgery for colorectal cancer using quality care indicators. *Can J Surg*. 2010; 53(4):232-240.
93. Messenger DE, Cohen Z, Kirsch R et al. Favorable pathologic and long-term

- outcomes from the conventional approach to abdominoperineal resection. *Dis Colon Rectum*. 2011; 54:793-802.
94. Widdifield J, Bombardier C, Bernatsky S et al. An administrative data validation study of the accuracy of algorithms for identifying rheumatoid arthritis: the influence of the reference standard on algorithm performance. *BMC Musculoskelet Disord*. 2014; 15:216-225.
95. Lee DS, Stitt A, Wang X et al. Administrative hospitalization database validation of cardiac procedure codes. *Med Care*. 2013; 51(4):e22-26.
96. Goldsbury DE, Armstrong K, Simonella L, Armstrong BK, O'Connell DL. Using administrative health data to describe colorectal and lung cancer in New South Wales, Australia: a validation study. *BMC Health Services Research*. 2012; 12:387-396.
97. Li X, King C, deGara C, White J, Winget M. Validation of colorectal cancer surgery data from administrative data sources. *BMC Medical Research Methodology*. 2012; 12:97-104
98. Tucker TC, Howe HL, Weir HK. Certification for population-based cancer registries. *J Reg MGMT*. 1999; 26:24-27.
99. Juurlink D, Preyra C, Croxford R et al. Canadian Institute for Health Information Discharge Abstract Database: a validation study. Toronto: *Institute for Clinical Evaluative Sciences*; 2006.
100. Van Walraven C, Austin PC, Manuel D, Knoll G, Jennings A, Forster A. The usefulness of administrative databases for identifying disease cohorts is increased with a multivariate model. *J Clin Epidemiol*. 2010; 63(12):1332-1341.

Table 1. Inclusion and exclusion criteria used to define rectal cancer cohort

1. Inclusion Criteria

- Diagnosis of Rectal Cancer (defined as distal tumor margin within 15cm of anal verge).
 - Age \geq 18 years
 - Surgical Resection of Rectal Cancer with Intent to Cure at TOH between 1996 and 2010
 - Pathology reports, hospitalization records and CAC notes must be available.
-

2. Exclusion Criteria

- Cancer of the Colon (Based on pathology reports and clinical notes).
 - Rectal cancer resections without curative intent (i.e. palliative resections).
 - Age \leq 18 years
-

Table 2. First round results of text-search algorithm* development. Each iteration was performed on a population of 694 known rectal cancer resections performed at The Ottawa Hospital.

<i>Iteration #</i>	<i>Key Phrases Leading to Study Inclusion**</i>	<i>Key Phrases Leading to Study Exclusion††</i>	<i>Test Cohort Reports Successfully Identified, N(%)</i>
1	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy")	None	694 (100)
2	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Endo" or "Endoscopy")	652 (93.9)
3 [^]	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	None	694 (100)
4	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Benign")	662 (95.3)
5	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Biopsy")	557 (80.3)
6	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Biopsy") within 100 spaces of ("Diagnosis")	612 (88.2)
7	("Rectum" or "Rectal") AND ("Cancer" or "Malignancy") within 20 spaces of ("Diagnosis" or "Diagnoses")	("Biopsy") within 20 spaces of ("Diagnosis")	628 (90.5)

* *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 next round of text-search algorithm development**

Table 3. Second round results of text-search algorithm* development. 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, non-colorectal cancer cases.

<i>Iteration #</i>	<i>Key Phrases Leading to Study Inclusion**</i>	<i>Key Phrases Leading to Study Exclusion††</i>	<i>Sensitivity (%)</i>	<i>Specificity (%)</i>
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.

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)

[^] **Combination of key words/phrases used for the final text-search algorithm**

Table 4. 2x2 table demonstrating sensitivity, specificity, PPV and NPV of a text-search algorithm applied to pathology reports for rectal cancer resections in OHDW (N=284,032).

	<i>True Rectal Cancer Resections*</i>		<i>Totals</i>
	+	-	
<i>Text Search Identified Rectal Cancer Resections~</i>	A 1038	B 4550	5588
	C 0 [†]	D 278,444	278,444
<i>Totals</i>	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

Table 5. Demographics for entire rectal cancer cohort created in Objective 1 (N=1038).

Age in years, mean (SD)	65.1 (12.2)
Female, N(%)	342 (33.0)
Type of Surgery, N(%)	
APR	351 (33.8)
LAR	596 (57.4%)
Hartmann	91 (8.8%)
Pathologic Stage	
0	89 (8.6)
1	320 (30.8)
2	251 (24.2)
3	327 (31.5)
4	51 (4.9)
Neoadjuvant Chemoradiation, N(%)	495 (47.7)
MVR, N(%)	74 (7.1)
Tumor Location*, N(%)	
Low	369 (35.6)
Mid	433 (41.7)
High	217 (20.9)
Surgical Technique, N(%)	
Laparoscopic	216 (20.8)
Open	790 (76.1)
Converted	32 (3.1)
Emergency Surgery, N(%)	27 (2.6)

* Tumor location is determined by its distance from the anal verge (Low = 0-5cm, Mid=6-10cm, High>10cm). Note that data for tumor distance from the anal verge are missing for 19 patients.
APR = Abdominoperineal Resection
LAR = Low Anterior Resection;
MVR = Multivisceral Resection

Table 6. Trends for surgical quality indicators (¶) for rectal cancer surgery at The Ottawa Hospital between January 1st, 2000 and December 31st, 2010 (N=972).

	2000 (N=33)	2001 (N=45)	2002 (N=67)	2003 (N=86)	2004 (N=88)	2005 (N=99)	2006 (N=118)	2007 (N=93)	2008 (N=116)	2009 (N=122)	2010 (N=105)	p- value*
Positive CRM, N(%)	1 (3.0)	10 (22.2)	4 (6.0)	6 (7.0)	3 (3.4)	5 (5.1)	10 (8.5)	5 (5.4)	7 (6.0)	12 (9.8)	3 (2.9)	0.24
Successful LN Retrieval, N(%)	10 (30.3)	21 (46.7)	23 (34.3)	37 (43.0)	58 (65.9)	50 (50.5)	71 (60.2)	59 (62.4)	73 (62.9)	99 (81.5)	92 (87.6)	<0.001
Positive Distal Margin, n(%)	0 (0.0)	1 (2.2)	0 (0.0)	1 (1.2)	0 (0.0)	0 (0.0)	0 (0.0)	1 (1.1)	3 (2.6)	0 (0.0)	0 (0.0)	0.97
APR's, n(%)	13 (39.4)	14 (31.1)	23 (34.3)	23 (26.7)	23 (26.1)	34 (34.3)	39 (33.1)	39 (41.9)	32 (27.6)	48 (39.3)	40 (38.1)	0.18

* Statistically significant trends of categorical variables determined using Cochran-Armitage Test for Trend

CRM = Circumferential Radial Margin

LN = Lymph Node

APR = Abdominoperineal Resection

¶ A successful outcome for each surgical quality indicator is defined as the following:

CRM: a negative margin (i.e. a positive CRM represents an unsuccessful outcome)

LN Retrieval: The retrieval of 12 or more lymph nodes in the surgical specimen

DRM: a negative margin (i.e. a positive DRM represents an unsuccessful outcome)

APR: Non-APR resections are considered a success (i.e. each APR is considered an unsuccessful outcome)

Table 7. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome *Circumferential Resection Margin (CRM)*[¶].

<i>Characteristic</i>	<i>Adjusted OR OR (95% CI)</i>	<i>p-value*</i>
Age	1.01 (0.99 – 1.04)	0.34
Female (ref = male)	0.76 (0.41 – 1.45)	0.38
Neoadjuvant (ref = no Neoadjuvant)	1.06 (0.55 – 2.04)	0.86
MVR (ref = no MVR)	2.59 (1.24 – 5.40)	0.01*
Tumor Stage (ref = 1)		
2	28.39 (3.73 – 215.95)	<0.01*
3	31.01 (4.16 – 231.13)	
4	111.37 (10.00 - >999.99)	
Emergency Surgery (ref = Elective)	10.11 (2.98 – 34.33)	<0.01*
Tumor Distance from anal verge in cm	0.88 (0.82 – 0.96)	<0.01*
Mode of Surgery (ref = Lap)		
Open	10.53 (1.40 – 78.98)	0.04*
Converted	26.70 (2.71 – 262.58)	
Presence of Metastases (ref = No Metastases)	0.88 (0.26 – 2.96)	0.84
Year	1.00 (0.92 – 1.09)	0.94

OR = Odds Ratio, CI = Confidence Interval

Neoadjuvant = Patient received preoperative pelvic radiation +/- chemotherapy

MVR = Multivisceral Resection

Converted = Intraoperative conversion from laparoscopic to open surgery

¶ = An event in this regression model represents a positive CRM, which is a failure of surgical quality (A positive margin means not all tumor cells were resected).

** = statistically significant test association*

Table 8. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome *Successful Lymph Node Retrieval*.

<i>Characteristic</i>	<i>Adjusted OR OR (95% CI)</i>	<i>p-value*</i>
Age	0.99 (0.98 – 1.01)	0.92
Female (ref = male)	0.86 (0.64 – 0.80)	0.30
Neoadjuvant (ref = no Neoadjuvant)	0.58 (0.42 – 0.80)	<0.01*
MVR (ref = no MVR)	1.14 (0.65 – 1.98)	0.65
Tumor Stage (ref = 1)		
2	1.53 (1.04 – 2.24)	0.03*
3	1.71 (1.20 – 2.43)	
4	1.10 (0.40 – 3.00)	
Emergency Surgery (ref = Elective)	0.94 (0.36 – 2.42)	0.89
Tumor Distance	1.04 (1.00 – 1.08)	0.04*
Mode of Surgery (ref = Lap)		
Open	0.97 (0.66 – 1.40)	0.97
Converted	1.09 (0.45 – 2.63)	
Presence of Metastases (ref = No Metastases)	2.04 (0.98 – 4.25)	0.06
Year	1.27 (1.21 – 1.33)	<0.01*

OR = Odds Ratio, CI = Confidence Interval

Neoadjuvant = Patient received preoperative pelvic radiation +/- chemotherapy

MVR = Multivisceral Resection

Converted = Intraoperative conversion from laparoscopic to open surgery

** = statistically significant test association*

Table 9. Results of multivariate logistic regression analysis examining both categorical and continuous variable predictors of the surgical quality outcome *Abdominal Perineal Resection*.

<i>Characteristic</i>	<i>Adjusted OR OR (95% CI)</i>	<i>p-value*</i>
Age	1.02 (1.00 – 1.03)	0.02*
Female (ref = male)	0.56 (0.35 – 0.91)	0.02*
Neoadjuvant (ref = no Neoadjuvant)	0.71 (0.44 – 1.15)	0.16
MVR (ref = no MVR)	2.48 (1.01 – 6.08)	0.04*
Tumor Stage (ref = 1)		
2	1.20 (0.66 – 2.20)	0.92
3	1.07 (0.62 – 1.85)	
4	1.96 (0.45 – 8.48)	
Emergency Surgery (ref = Elective)	0.94 (0.36 – 2.42)	0.98
Tumor Distance	0.45 (0.41 – 0.50)	<0.01*
Mode of Surgery (ref = Lap)		
Open	0.72 (0.40 – 1.29)	0.58
Converted	0.43 (0.10 – 1.84)	
Presence of Metastases (ref = No Metastases)	0.80 (0.27 – 2.40)	0.69
Year	0.97 (0.91 – 1.04)	0.40

OR = Odds Ratio, CI = Confidence Interval

Neoadjuvant = Patient received preoperative pelvic radiation +/- chemotherapy

MVR = Multivisceral Resection

Converted = Intraoperative conversion from laparoscopic to open surgery

** = statistically significant test association*

Table 10. 2x2 table demonstrating sensitivity, specificity, PPV and NPV of previously published coding algorithm for identifying rectal cancer resections in ICES when compared to a validation cohort (N=664,075).

	<i>True Rectal Cancer Resections*</i>		<i>Totals</i>
	+	-	
<i>ICES Identified Rectal Cancer Resections~</i>	A 735	B 396	1131
	C 86	D 662,858	662,944
<i>Totals</i>	821	663,254	664,075

Sensitivity = $A/A+C = 735/821 = 89.5\%$

Specificity = $D/B+D = 662,858/663,254 = 99.9\%$

PPV = $A/A+B = 735/1131 = 64.9\%$

NPV = $D/C+D = 662,858/662,944 = 99.9\%$

*True Rectal Cancer Resections refer to the validated cohort of rectal cancer patients validated through manual chart review in Objective #1 of the present study.

~ ICES Identified Rectal Cancer Resections refers to the rectal cancer resections identified through previously published coding algorithms mining ICES databases

PPV = Positive Predictive Value

NPV = Negative Predictive Value

ICES = Institute for Clinical and Evaluative Sciences

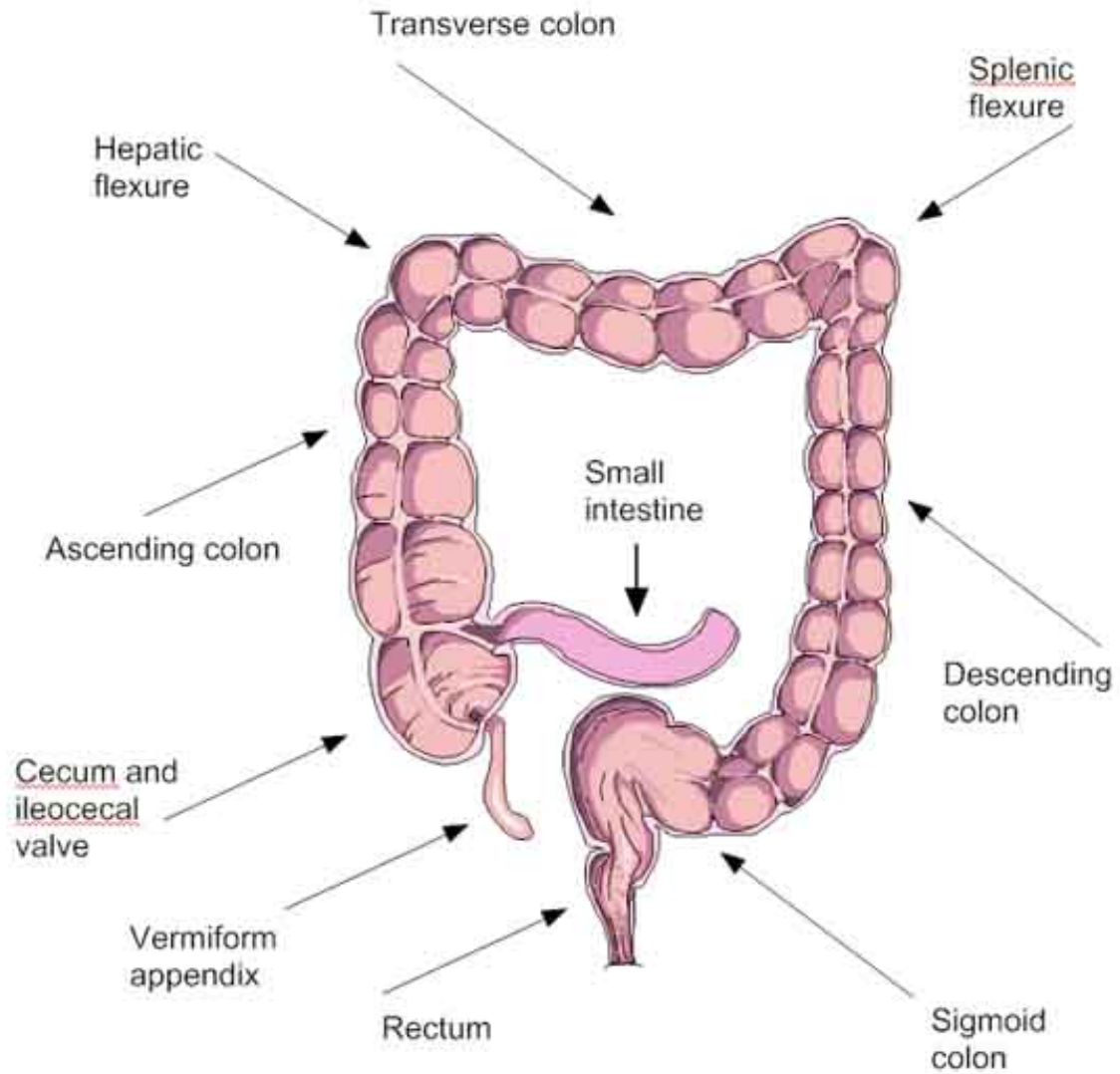


Figure 1. Anatomical divisions of the colon and rectum (from <http://jonbarron.org/article/colon#.VPet-iIDbd4>).

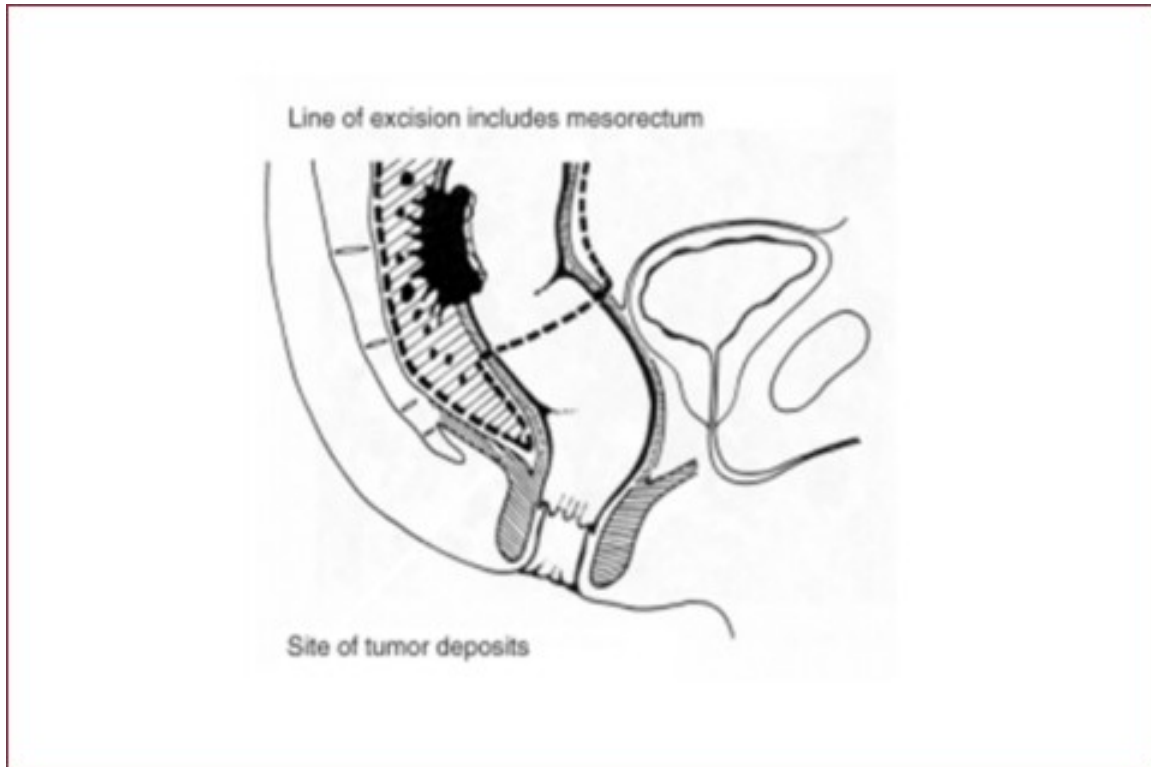


Figure 2. Graphical representation of mesorectal fascial planes followed during Total Mesorectal Excision (TME) (from Heald et al., 1982).

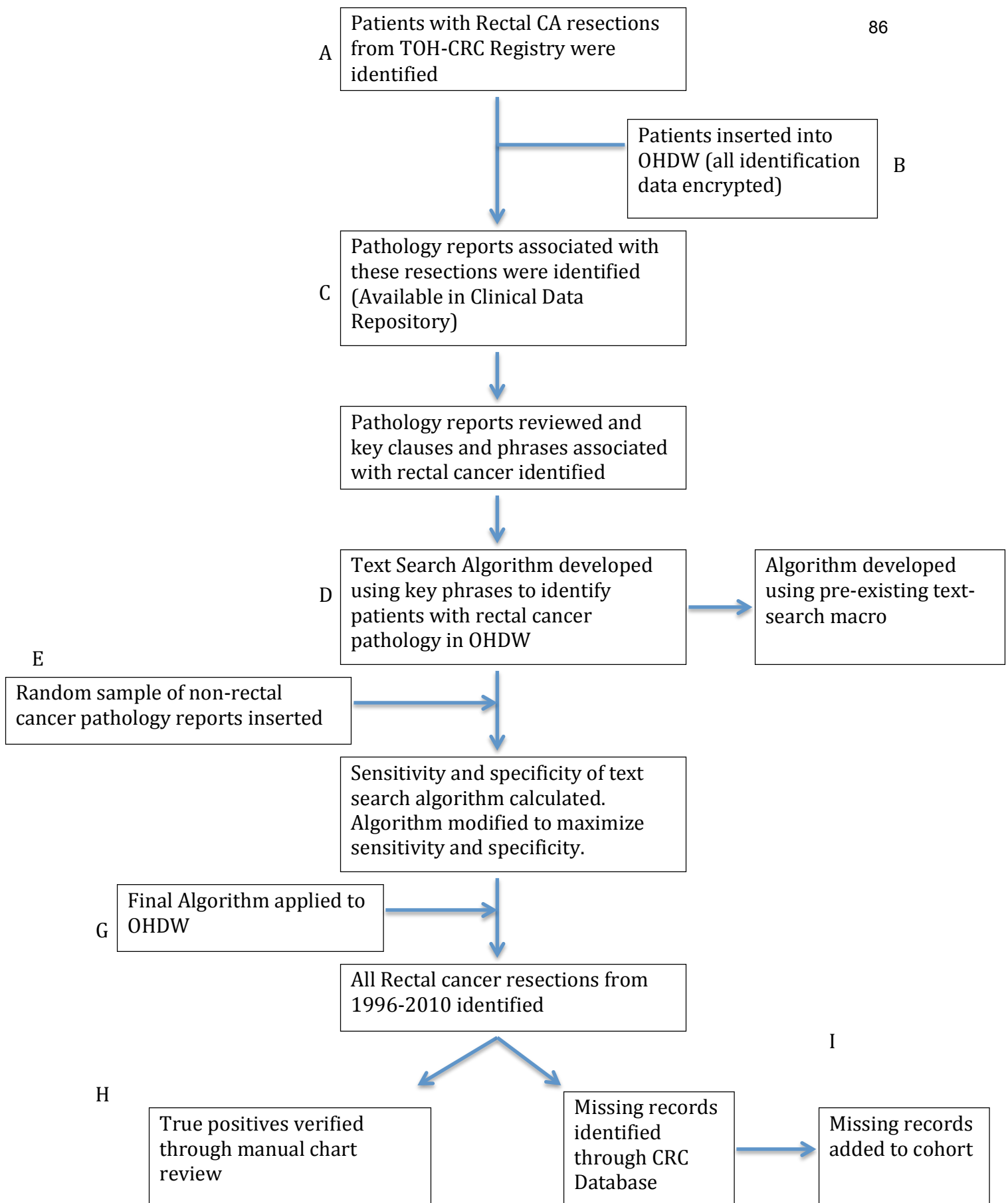


Figure 3. Flow diagram depicting methodology used to complete Objective 1 (Identification and validation of rectal cancer cohort) Letters A-G represent specific steps referred to in text. The final cohort of rectal cancer patients was created from the combination of box H and I (Figure 4).

TOH-CRC = The Ottawa Hospital - Colorectal Cancer; OHDW = Ottawa Hospital Data Warehouse

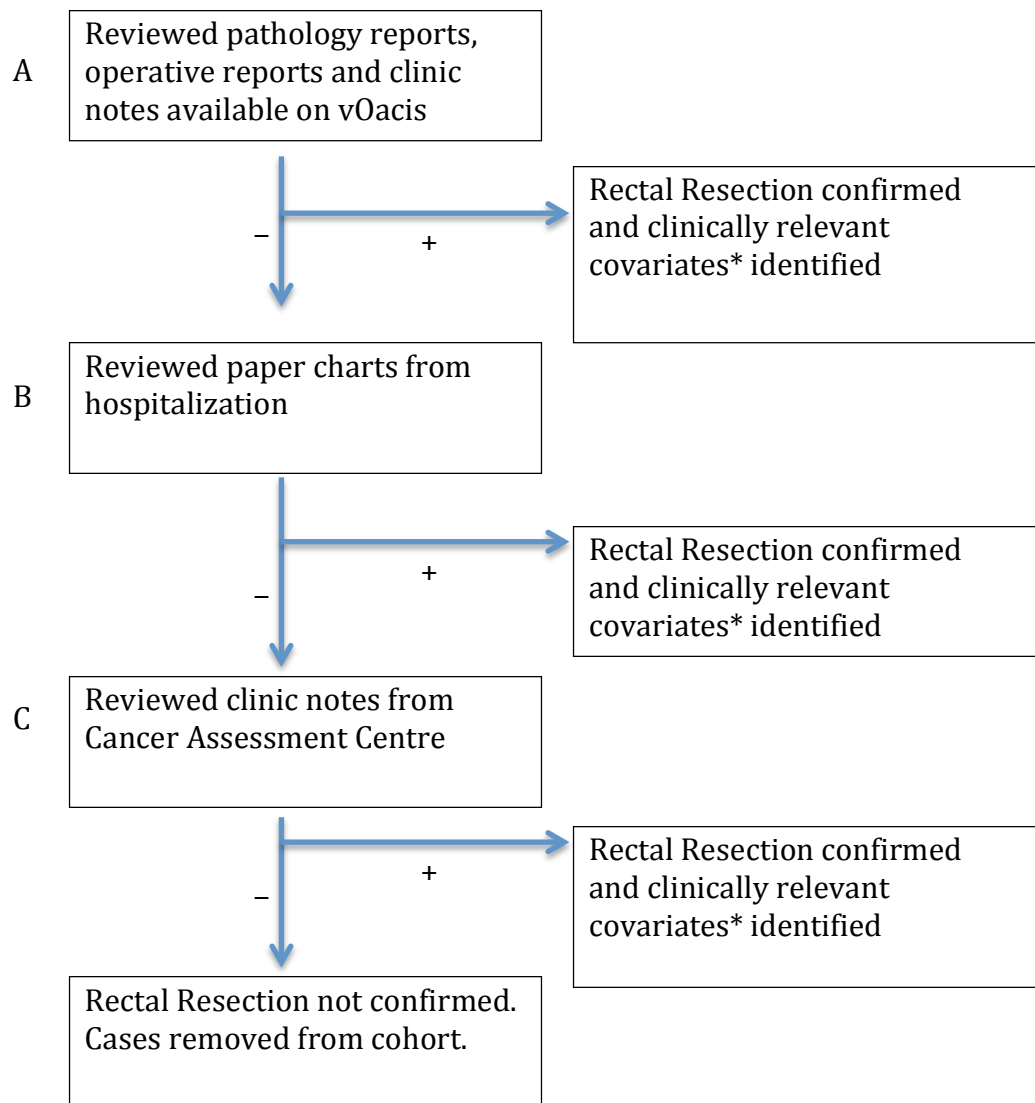


Figure 4. Flow diagram depicting methodology used for manual chart review to confirm diagnosis of rectal cancer and identify relevant covariates

*Clinically relevant covariates refer to variables that need to be considered when assessing outcomes in rectal cancer. Examples include stage of cancer at time of resection, age of patient, and exposure to pre-operative chemotherapy and radiation. All covariates were determined by clinical supervisors who are experts in the field of colon and rectal cancer surgery and are outlined in detail in the text of this paper.

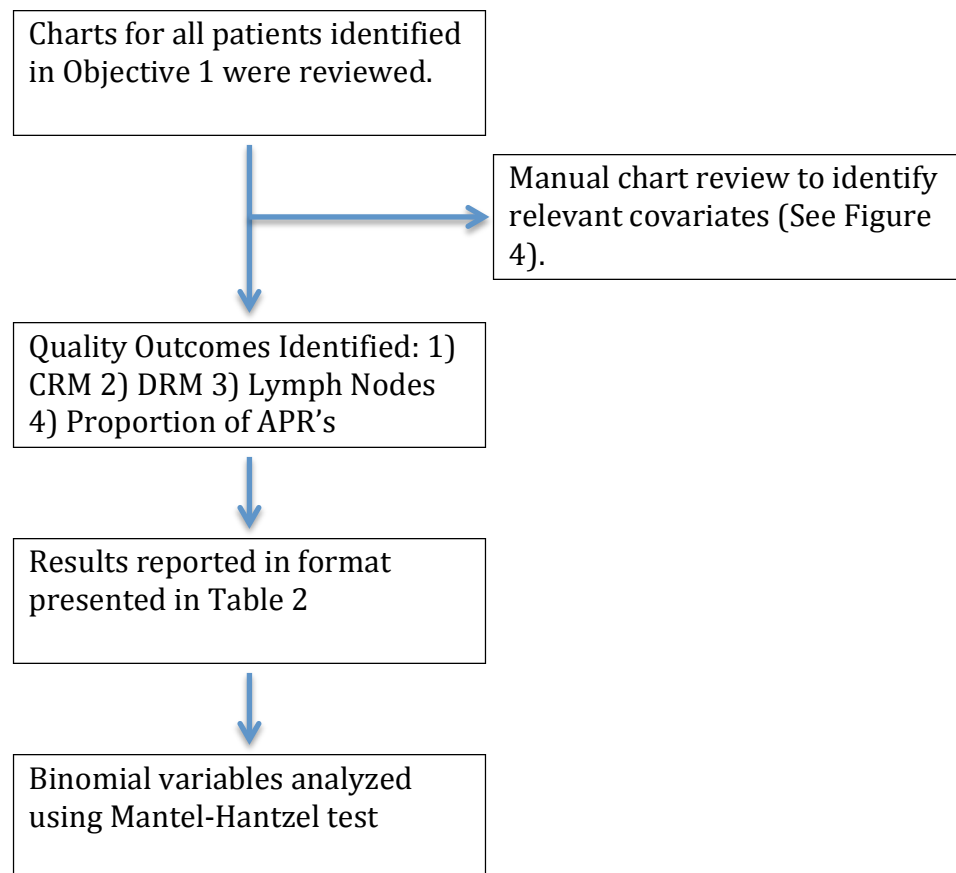


Figure 5. Flow diagram depicting methodology used to complete Objective 2 (assessment of surgical quality indicators)

CRM = Circumferential Resection Margin

DRM = Distal Resection Margin

APR = Abdominoperineal Resection

Lymph Nodes - Refers to the proportion of resections where ≥ 12 nodes are reported in the pathology report

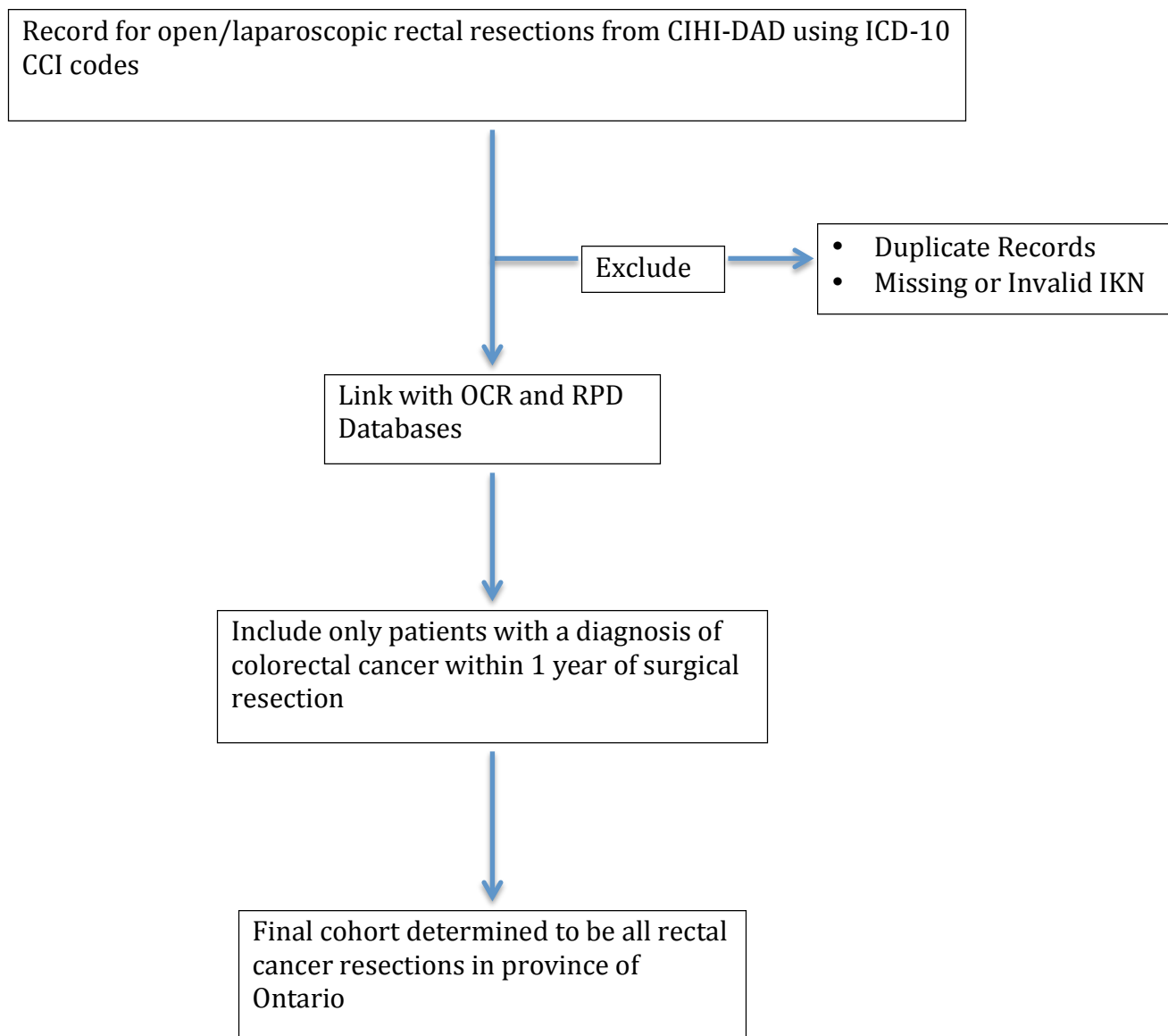


Figure 6. Flow diagram illustrating methodology used to identify rectal cancer resections through administrative databases in previously published studies (re-created with permission from the authors^{39,40})

CIHI-DAD = Canadian Institute for Health Information, Discharge Abstract Database

ICD-9 = International Classification of Disease, version 9

CCI = Canadian Classification of Health Interventions

OCR = Ontario Cancer Registry

RPD = Registered Persons Database

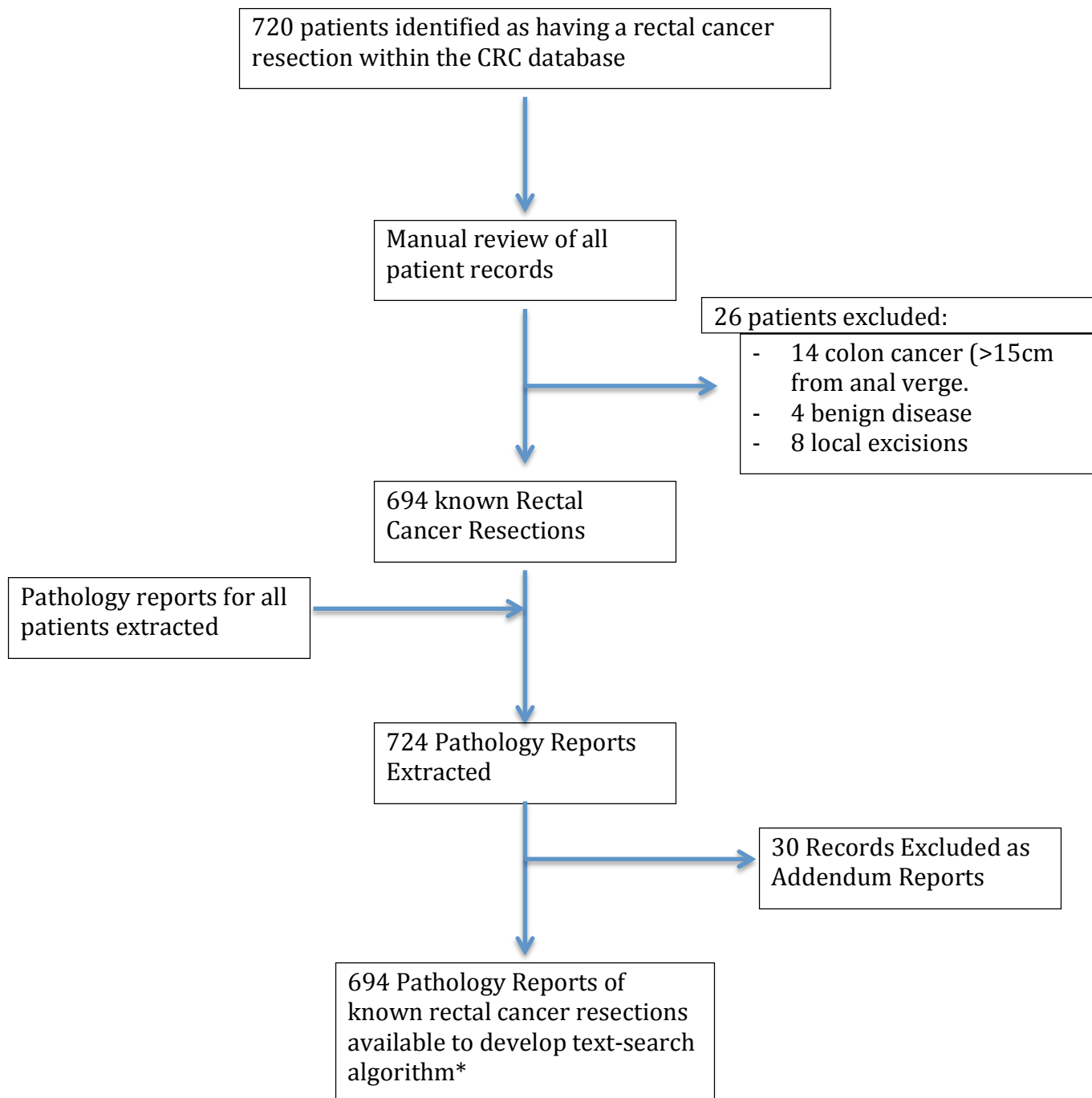


Figure 7. Flow Diagram illustrating methods used to identify pathology reports of known rectal cancer resections at The Ottawa Hospital from 2002 to 2010 (*This group of reports was later used to develop a text-search algorithm for accurately identifying rectal cancer resections through administrative databases).

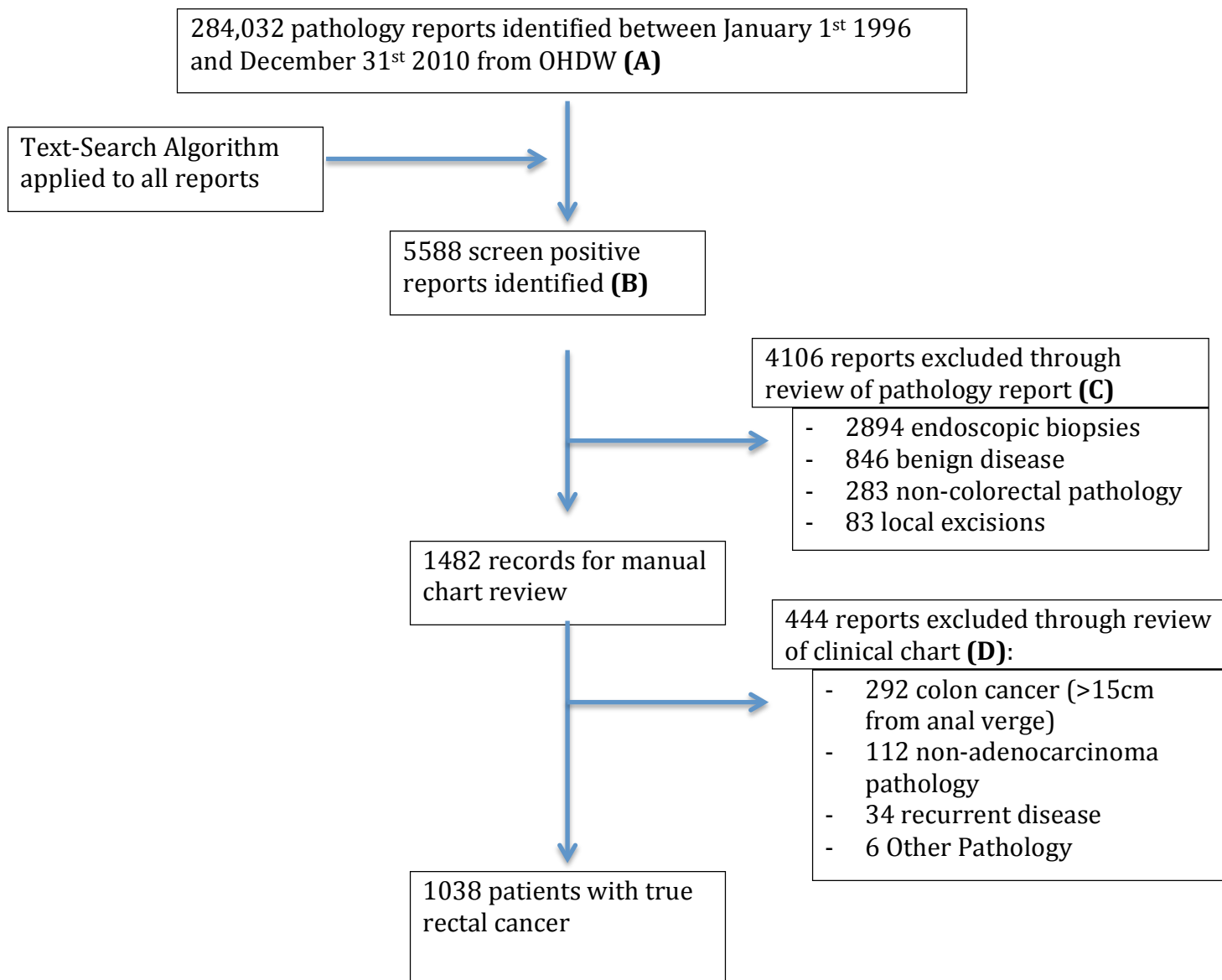


Figure 8. Flow Diagram illustrating identification and validation of all rectal cancer resections at The Ottawa Hospital between January 1st, 1996 and December 31st, 2010.

OHDW = Ottawa Hospital Data Warehouse

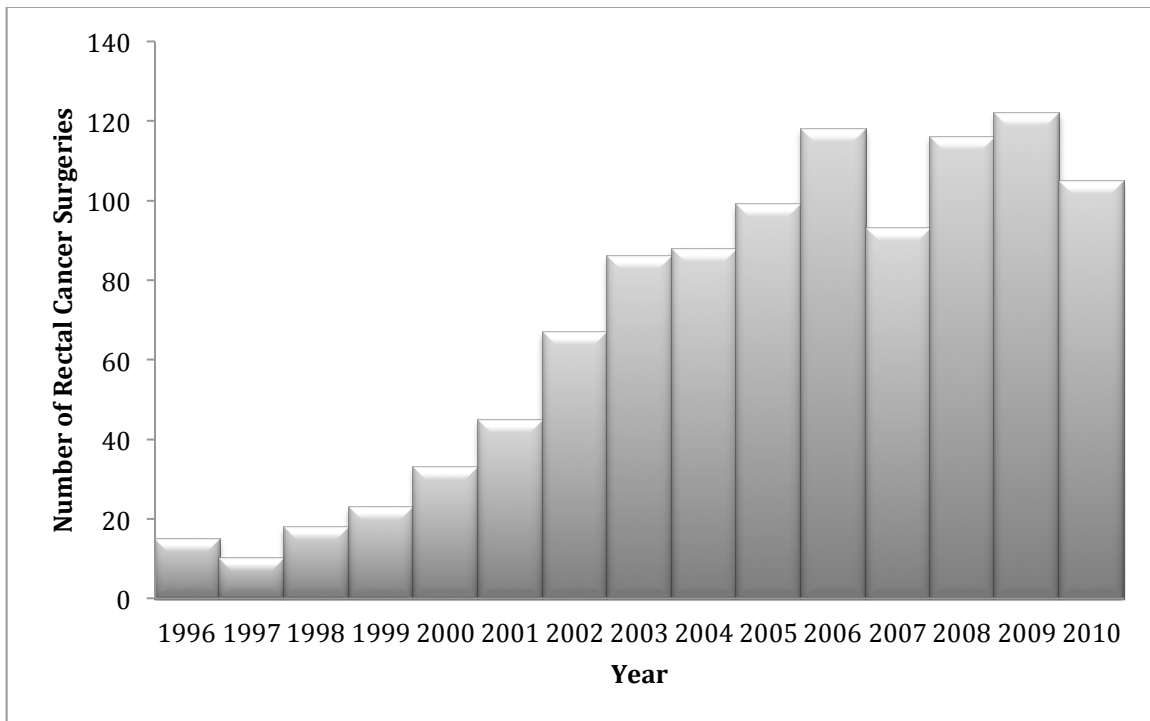


Figure 9. Total number of rectal cancer resections performed per year at The Ottawa Hospital between 1996 and 2010 (inclusive).

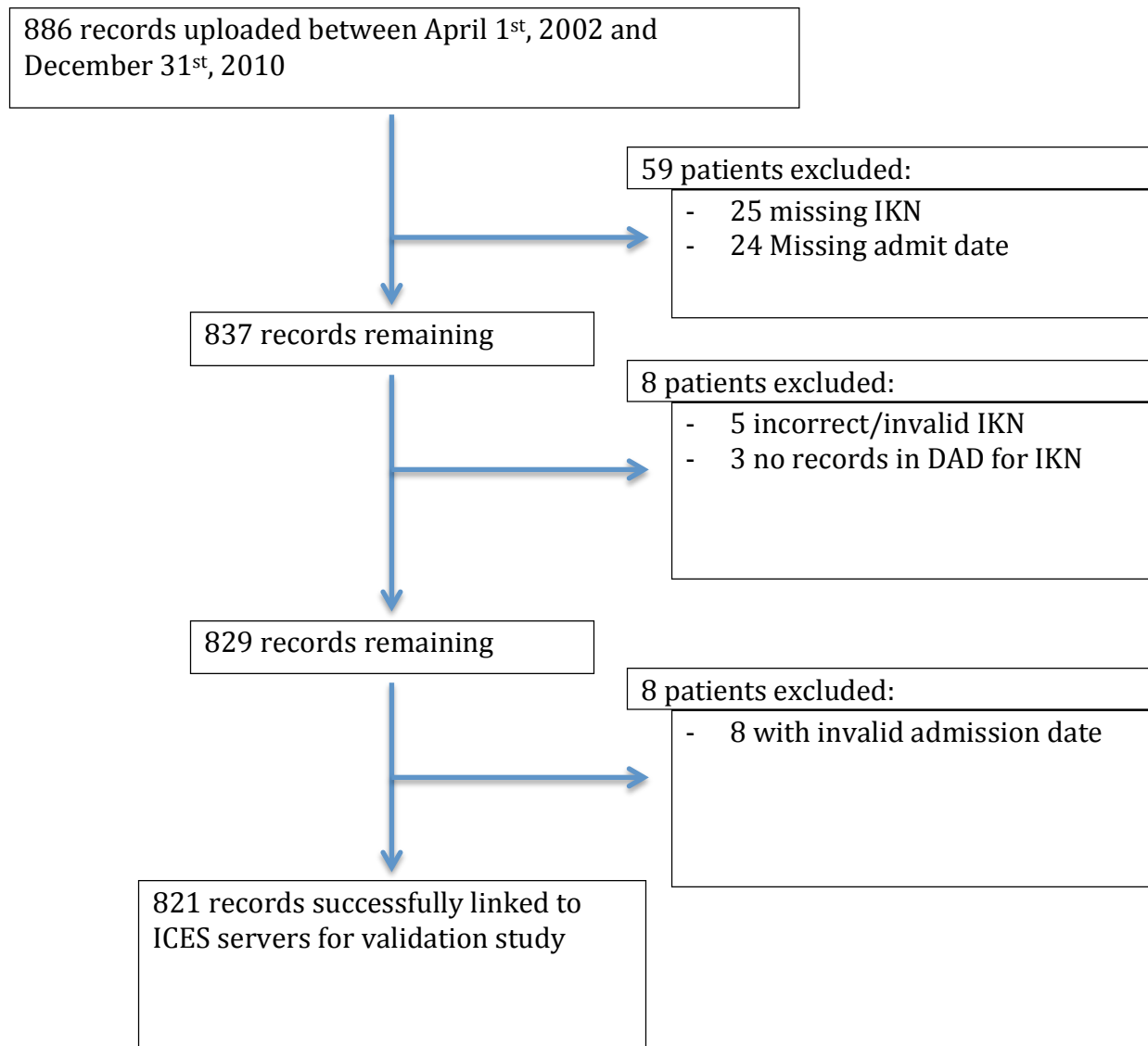


Figure 10. Flow Diagram illustrating patient records from validation cohort that were successfully uploaded to ICES databases.

IKN = ICES Key Numbers

DAD = Discharge Abstract Database

ICES = Institute for Clinical and Evaluative Sciences

Appendix 1: CCI* Codes Used to Define Cohort

Code	Description
1NQ87DF	Excision partial, rectum using laparoscopic approach with enterocolostomy
1NQ89GV	Excision total, rectum using combined laparoscopic with perineal approach with coloanal or ileoanal anastomosis
1NQ87PB	Excision partial, rectum using open approach with colorectal anastomosis
1NQ87RD	Excision partial, rectum using open approach with colorectal or coloanal anastomosis
1NQ89KZ	Excision total, rectum using open abdominoperineal approach
1NQ89KZXXG	Excision total, rectum using abdominoperineal approach pouch formation with ileoanal anastomosis
1NQ89SF	Excision total, rectum using open approach with bypass technique
1NQ89SFXXG	Excision total, rectum abdominal approach and pouch formation
1NQ90LAXXG	Excision total with reconstruction, rectum using open approach with ileoanal anastomosis
1NQ87DE	Excision partial, rectum using laparoscopic approach with colorectal anastomosis
1NQ89LH	Excision total, rectum using open abdominoperineal approach
1NQ89LHXXG	Excision total, rectum using abdominoperineal approach and continent ileostomy formation
1NQ89RS	Excision total, rectum using open anterior approach with terminal colostomy
1NQ89RSXXG	Excision total, rectum using abdominal approach and continent ileostomy formation
1NQ89AB	Excision total, rectum using combined approach, laparoscopic, stoma formation with distal closure
1NQ87DX	Excision partial, rectum using laparoscopic approach, stoma formation and distal closure
1NQ87TF	Excision partial, rectum using open approach, stoma formation and distal closure

*CCI = Canadian Classification of health Interventions