

A Data Management Approach to Quality Assurance Using Colorectal Carcinoma Reports From Two Institutions as a Model

John P. Sorge, MD,¹ C. Reid Harmon, PhD,² Susan M. Sherman, PhD,³ and E. Eugene Baillie, MD^{4*}

Key Words: Quality improvement; Automatic; Lymph node; Colorectal carcinoma

DOI: 10.1309/330FJ8WBF4E52A50

Abstract

We used data management software to compare pathology report data concerning regional lymph node sampling for colorectal carcinoma from 2 institutions using different dissection methods. Data were retrieved from 2 disparate anatomic pathology information systems for all cases of colorectal carcinoma in 2003 involving the ascending and descending colon. Initial sorting of the data included overall lymph node recovery to assess differences between the dissection methods at the 2 institutions. Additional segregation of the data was used to challenge the application's capability of accurately addressing the complexity of the process. This software approach can be used to evaluate data from disparate computer systems, and we demonstrate how an automated function can enable institutions to compare internal pathologic assessment processes and the results of those comparisons. The use of this process has future implications for pathology quality assurance in other areas.

The 1999 College of American Pathologists Consensus Statement lists regional lymph node metastases as a category 1 prognostic factor in conjunction with local extent of tumor assessed by pathologic examination.¹ The recommendations for the retrieval of a minimum number of lymph nodes in colorectal carcinoma have varied from 6 to 20.²⁻⁷ The College of American Pathologists Working Group recommended that at least 12 regional lymph nodes be obtained from a radical colorectal resection.⁸ The American Joint Commission on Cancer's *Cancer Staging Manual* recommends retrieval of 7 to 14 lymph nodes with a proviso that in patients treated with preoperative radiation, fewer recovered lymph nodes may be acceptable.⁹ The thoroughness of the pathologic assessment is critical in staging, but other factors might influence nodal sampling in colorectal carcinoma, including surgical and patient factors. For example, specimens from resections of the ascending colon generally are longer than those from the descending colon and, consequently, yield more lymph nodes.^{10,11}

The approach in the present study was to evaluate specimen data from 2 disparate anatomic pathology information systems, in which one institution (institution A) used a standard manual method of dissection and the other institution (institution B) used a fat clearance method. An automated proprietary software method bridged information processing from the separate anatomic pathology information systems and maintained patient confidentiality. This automated application is expected to facilitate process improvement and foster quality assurance by offering lymph node recovery information for comparison between the 2 institutions and to generate other process comparisons for noncolorectal carcinoma specimens. Future considerations for a more global software

approach to quality assurance include adding other institutions and disparate computer systems.

Materials and Methods

Data Management Process

By using proprietary computer-based data mining software (inREACH, Anderson, SC), we electronically obtained pathology reports from 2 sites with disparate computer systems. Colorectal carcinoma cases were selected to prove and validate the model. The colorectal carcinoma cases also were manually collated for the entire calendar year 2003 by both institutions to validate the automated process.

The data management software selects pathology reports within a day of being signed out, and it identifies the source fields of the report (eg, gross, microscopic, final diagnosis). An independent module then identifies colorectal carcinoma cases and determines the number of lymph nodes and other derivative fields (specimen type, specimen location, length of the colon portion of specimen, tumor size, nodal status, and metastatic status) through pattern recognition. The patterns were built by studying hundreds of cases to determine the ways in which information was supplied by the signing pathologist. All colorectal carcinoma cases from 2003 were compared with the output of the classification process to carefully refine the patterns to be as specific as possible without yielding false-positive matches. The software can make decisions only about the data present in the report, so there will always be occasional cases requiring manual review (eg, words referring to other organs such as ovary or stomach would cause the case to be classified as indeterminate and require manual review).

At study inception, pathology cases from institution B could be obtained only by subscribing to a Health Level 7 (Health Level Seven, Ann Arbor, MI) feed from its CoPath Plus anatomic pathology information system (Cerner, Kansas City, KS). Therefore, a personal computer was used at institution B to receive the Health Level 7 data and transform the data stream into coherent reports, complete with source field annotation. These reports were bundled together daily and sent in a single file to a central server. For secure communication, privacy encryption (PGP [pretty good privacy]) using public and private keys was used. First, the outgoing file was signed with institution B's private key, then encrypted with a public key, and transmitted via standardized file download/upload (FTP [file transfer protocol]) to the central server. Because institution A uses a different anatomic pathology information system (Cerner Millennium), institution A's reports were generated via Cerner proprietary query language, CCL (Cerner Command Language). A daily morning process

obtained cases signed out or modified the previous day and annotated the source fields. The resulting file was signed with institution A's private key, encrypted with the public key, and transmitted via file transfer protocol to the central server.

The common format for transmission allows the commingling of pathology reports from multiple sites. Instead of relying on reports that have identical structure, extensible markup language is used as the transmission format, identifying not only individual reports, but also the fields themselves through the use of markup tags.

This format allows the central processing server to identify the source fields without being familiar with the structure of the source report. Classification of the derivative fields occurs next; the classification process applies patterns in the form of regular expressions to the source fields of interest to find the derivative fields. For example, the number of lymph nodes may be identified by a statement in the diagnosis such as: "Four of eleven lymph nodes showed metastasis." This statement reveals the number of lymph nodes and the number that were metastatic. The statement might also be written: "Metastasis identified in the lymph nodes (4/11)." The pattern recognition is trained to look for numbers in conjunction with the terms "lymph nodes" and "metastasis," as well as similar forms (eg, "no metastatic tumor"). Sometimes one or more derivative fields are implicit; for example, "Seven lymph nodes recovered, all benign." In this case, the number of lymph nodes is given, but the number that was metastatic, 0, is implied by the term benign. Performing the classification at a central location allows for changes and additions in the recognized patterns more easily. Even once-processed reports can be reprocessed when the set of patterns is changed.

After the derivative fields are classified in the report files, the output files are sent to a database-input process. A Microsoft structured query database (MS-SQL; Microsoft, Redmond, WA) was used for its high performance/cost ratio. The database-input process updates the database by reading the report files and sending commands to the database in SQL language. The extensible markup language format of the files allows for the source and derivative tags to be mapped to tables, fields, and data types in the database. This makes it possible to add a new derivative field in the classification process without manually changing the database or the database-input process; the file itself indicates which fields should be stored and where.

All of these processes are automated: data are collected at the source sites, gathered into files daily, transmitted securely, classified, and input into the database, all by separate processes launched automatically by the respective systems. Fail-safe processes were included to prevent loss of data. Finally, the central database server also serves a Web site, which allows authenticated access to the reports stored in the database for quality assurance purposes. Security is arranged such that each report is

marked with a source site, and each user is also marked with a source site. A user can view all information on cases from his or her own site, but only nonsensitive (ie, stripped of demographic data) information on cases from other sites. Fields are marked sensitive by default and can be released only as nonsensitive by system administrators. Examples of sensitive information include the patient's name, medical record number, source site, case number, and sign-out pathologist. In addition to the ability to search for groups of cases matching specific criteria, the Web site offers statistical studies, such as summation of the average number of lymph nodes recovered by location. This capability makes it possible for pathologists to perform additional desired studies for quality assurance.

Specimen Management Process

The ascending and descending colon resection specimens from institution B were submitted fresh for fat clearance and were evaluated manually to initially assess tumor size, distance to proximal and distal margins, extent of tumor penetration, and radial margin status. The fat was stripped from the bowel segment and serially sectioned at 0.5- to 1.0-cm intervals; an intact mesenteric fat border was maintained. The fat was placed in a separate container submerged in a formaldehyde, glacial acetic acid, ethyl alcohol, and deionized water solution (Dissect Aid, Decal, Tallman, NY) for 2 to 3 hours before dissection. The ascending and descending colon resection specimens from institution A were submitted in formalin, and manual dissection of the pericolic adipose tissue was performed with the adipose tissue attached to the segment of colon. The dissection was completed on the same day as the specimen was received. At both sites, representative sections of grossly involved lymph nodes were taken, whereas grossly uninvolved lymph nodes were submitted in their entirety.

Statistical Methods

A 2-way analysis of variance was used to evaluate possible differences in the square root-transformed number of nodes across the 2 sites (A and B) and the 2 sides (descending and ascending). The Levene test was used to test the assumption of equal variances between groups. Square root transformations successfully normalized the distribution of nodes. The nonparametric χ^2 analysis was used to study differences among categorically distributed data. Data were validated using alternate transformations and analyses with comparable results.

Results

The databases for institutions A and B were searched for colorectal carcinoma cases meeting the criteria of a colectomy as the specimen type and adenocarcinoma identified in the diagnosis. The search language segregated cases into the

ascending, transverse, and descending colon. Cases from the transverse colon were eliminated from the study owing to insufficient sample size. The data were confirmed manually before implementing the automated method of data processing, along with ongoing quality control measures.

A total of 211 specimens representing 2,368 lymph nodes (mean, 11.2 lymph nodes per specimen) from the 2 institutions were retrieved from the database. **Table 1** gives lymph node recovery data for the 2 institutions. The difference in the number of lymph nodes harvested at the 2 institutions was statistically significant. Breakpoints for lymph node recovery were segregated into groups representing 0 to 10, 11 to 20, and greater than 20 per specimen. **Table 2** (2×3 contingency table) gives the number of cases at each institution in these categories. Institution B, with the fat clearance method, had significantly fewer cases with fewer than 11 lymph nodes per specimen and significantly more cases with more than 20 lymph nodes per specimen ($P < .05$; independent samples χ^2 test).

The data were sorted further by specimen location and whether lymph node metastases were present or absent. **Table 3** gives data from the 2 institutions according to specimen location (ascending vs descending colon) and mean lengths for colon segments from the right and left sides. A statistical difference was observed in lymph node recovery from the ascending colon compared with the descending colon at institution B but not at institution A. When data were combined from both institutions by specimen location, a significant difference was maintained between lymph nodes harvested from the ascending vs descending colon. **Table 4** gives the comparison of data from the 2 institutions by specimen location and whether the lymph nodes were uninvolved or involved by metastatic carcinoma. At institution A, 48 of 131 lymph nodes

Table 1
Lymph Node Recovery From Institutions A and B

	Institution A	Institution B	P
Total No. of specimens recovered	87	124	<.001
Total No. of lymph nodes harvested	765	1,603	<.001
Mean No. of lymph nodes per specimen	8.8	12.9	—

Table 2
Observed and Expected Frequencies for Number of Lymph Nodes From a 2×3 Contingency Table*

No. of Lymph Nodes	Institution A (n = 87)	Institution B (n = 124)	Expected Frequency (%)
<11	53 (61)	53 (43)	52
11-20	34 (39)	51 (41)	40
>20	0 (0)	20 (16)	8

* Data are given as number (percentage) unless otherwise indicated.

Table 3
Lymph Node Recovery by Specimen Location

	Ascending Colon	Descending Colon	P
Institution A			
No. of specimens recovered	42	45	>.05
No. of lymph nodes harvested	374	391	.951
Mean No. of lymph nodes per specimen	8.9	8.7	—
Mean colon segment length per specimen (cm)	21.7	17.8	<.001
Institution B			
No. of specimens recovered	59	65	<.001
No. of lymph nodes harvested	920	683	<.001
Mean No. of lymph nodes per specimen	15.6	10.5	—
Mean colon segment length per specimen (cm)	25.5	21.3	<.001
Totals from institutions A and B			
Total No. of specimens recovered	101	110	>.05
Total No. of lymph nodes harvested	1,294	1,074	.032
Mean No. of lymph nodes per specimen	12.8	9.8	—

from 13 specimens from the ascending colon and 64 of 174 lymph nodes from 19 specimens from the descending colon were involved by metastatic carcinoma. At institution B, 132 of 378 lymph nodes from 26 specimens from the ascending colon and 96 of 307 lymph nodes from 22 specimens from the descending colon were involved by metastatic carcinoma.

Cases with metastatic carcinoma constituted 37% (32/87) of the total cases at institution A and 38.7% (48/124) of the total cases at institution B. A χ^2 analysis did not demonstrate statistical significance between cases with and without metastatic disease at each of the institutions. An analysis of variance demonstrated statistical significance ($P = .004$) only for descending colon specimens at institution B between cases with and without metastases.

Discussion

Despite the fact that multiple non-pathology-related variables affect the number of lymph nodes in colorectal cancer specimens, fully accurate pathologic assessment is crucial to patient outcome and treatment and is desired by all. Sentinel lymph node mapping is not a standard of care in colorectal cancer as it is for melanoma and breast cancer. The detection

of very small micrometastatic deposits in lymph nodes in colon cancer specimens is very important by virtue of its N1 designation in the TNM classification system.

There are different approaches to improve evaluation of the colorectal regional lymph node basin. Multiple studies have addressed different methods of fat clearance with subsequent improvement in lymph node recovery.¹²⁻¹⁸ In addition, a method of improved histologic sectioning has been reported that has led to increased detection of lymph node metastases.¹⁹ Molecular techniques have been used to identify micrometastatic disease.^{20,21} In the present study, a fat-clearing solution of formaldehyde, glacial acetic acid, ethyl alcohol, and deionized water was used at one institution and compared with the standard manual method used at the second institution. Our data support the use of a fat-clearing solution to improve lymph node dissections of colorectal cancer specimens. This particular method is valuable because it can be incorporated into the daily workflow pattern without lengthening the turnaround time for reporting colorectal cancer cases. We recommend that to maximize the detection level, the stripped fat from the colon be placed and kept in the clearing agent for 2 to 3 hours, divided into 0.5- to 1.0-cm slices, and covered by a sufficient volume of the clearing agent.

Table 4
Lymph Node Recovery by Specimen Location and Presence or Absence of Metastatic Disease

	Ascending Colon		Descending Colon	
	Present	Absent	Present	Absent
Institution A				
No. of specimens recovered	13	29	19	26
No. of lymph nodes harvested	48/131	243	64/174	217
Mean No. of lymph nodes per specimen	10.1	8.4	9.2	8.3
Institution B				
No. of specimens recovered	26	33	22	43
No. of lymph nodes harvested	132/378	542	96/307	376
Mean No. of lymph nodes per specimen	14.5	16.4	14.0	8.7

The fat clearing method increases the yield of lymph nodes recovered from colorectal cancer specimens by facilitating the visual detection of smaller lymph nodes. The increased yield of lymph nodes seems to affect the prognosis by increasing the probability of detecting lymph node metastases.² In addition, the identification of small (ie, <5 mm) lymph nodes in itself is an important event in accurate staging.^{22,23} Another advantage of using this type of fat clearance is that the adipose tissue can be retained in this solution overnight for reharvesting the next day without decrement in the histologic features of the lymph nodes. Attention can be directed to the smallest of lymph nodes that were undetectable the previous day if the initial lymph node harvest is considered suboptimal.

Several reports have highlighted the difference in the number of lymph nodes retrieved from ascending vs descending colon specimens.^{10,11} The difference in lymph node recovery is attributed to the differences in length of specimens from the ascending descending colon. Colectomy specimens from the ascending colon generally are longer than descending counterparts, which is supported from our mean colon length measurements. Our data support a statistically significant difference in the number of lymph nodes recovered from ascending and descending colon specimens when data from both institutions were combined. However, a statistically significant difference was not found between ascending and descending colon specimens from institution A. At this point, we agree with a "minimum" number of recoverable lymph nodes from colorectal specimens but encourage continuing to separate data by specimen location. Ultimately, a different minimum number of recoverable lymph nodes might be necessary when evaluating a specimen from the ascending colon vs a specimen from the descending colon.

Although our number of cases is small compared with numbers in other studies,²⁴ the data are similar in the percentages of cases in which the number of lymph nodes retrieved was less than the recommended minimum of 12. In the present study, pretreatment of pericolic and mesenteric fat with a clearing agent improved lymph node retrieval. In addition, the number of cases in the higher lymph node recovery category (>20 lymph nodes) was greater using the fat clearance method. An immediate expectation in a process improvement program is to concentrate on reducing the number of cases that have lymph node recoveries less than the recommended minimum. However, there may be an irreducible number based on uncontrollable patient and surgical factors.

Recently, Wright et al²⁵ reviewed the issues surrounding suboptimal recovery of lymph nodes in colorectal cancer specimens. In their setting there exists an important role for an education strategy among pathologists and surgeons to improve the assessment of colorectal cancer specimens. The process presented in our study offers a novel approach to quality assurance among pathologists at 2 institutions using

disparate information systems for the purpose of comparing overall processes involved in assessment of colorectal cancer specimens. Before this study, participants were aware of national recommendations but had no reasonable way of comparing their performance with that of another institution. The timing of the study was opportune, because one institution used a fat clearing method and the other did not.

The data supported that process improvement might occur by incorporating fat pretreatment into the assessment of colorectal carcinoma specimens. There is a need for continued improvement owing to the considerable percentage of cases in the suboptimal recovery category of 0 to 10 lymph nodes per case. Ongoing efforts at both institutions are underway to decrease the percentages of cases for all categories of 0 to 10 lymph nodes per case. The automated approach for data collation provides a real-time method of bridging the barriers between disparate information systems, decreases the time and workforce necessary for manual data checking, and provides the foundation for an intensive quality assurance program at these 2 and potentially other institutions.

From the ¹Department of Pathology and ³Quality Programs, Rex Hospital, Raleigh, NC; ²inREACH, Anderson, SC; and ⁴Department of Pathology, Anderson Area Medical Center, Anderson, SC.

Supported in part by Pathology Services Associates, Florence, SC (partial funding of computer programming and partial computer server and data storage support), and a Rex Cancer Grant (funding for partial computer programming, computer server, and data storage support) from Rex Hospital.

Address reprint requests to Dr Sorge: Dept of Pathology, Rex Hospital, 4420 Lake Boone Trail, Raleigh, NC 27607.

Acknowledgments: We thank Tina Gowin, University of North Carolina at Chapel Hill, and Anna Butler, North Carolina State University, for data collation; Barbara Foreman, Department of Pathology, Rex Hospital, for data collation and medical transcription; Laurie K. Sorge, PhD, Raleigh, NC, for active participation in manuscript review; and Dominick T. Moore, Biostatistics in Comprehensive Cancer Department, Memorial Hospitals, University of North Carolina at Chapel Hill, for assistance with the review of statistical analysis.

** Dr Baillie is President of inREACH and was instrumental in the software application and pathologic assessment.*

References

- Compton CC, Fielding LP, Burgart LJ, et al. Prognostic factors in colorectal cancer: College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med.* 2000;124:979-994.
- Hernanz F, Revuelta S, Redondo C, et al. Colorectal adenocarcinoma: quality of the assessment of lymph node metastases. *Dis Colon Rectum.* 1994;37:373-377.
- Cianchi F, Palomba A, Boddi V, et al. Lymph node recovery from colorectal tumor specimens: recommendation for a minimum number of lymph nodes to be examined. *World J Surg.* 2002;26:384-389.

4. Swanson RS, Compton CC, Stewart AK, et al. The prognosis of T3N0 colon cancer is dependent on the number of lymph nodes examined. *Ann Surg Oncol*. 2003;10:65-71.
5. Tepper JE, O'Connell MJ, Niedzwiecki D, et al. Impact of number of nodes retrieved on outcome in patients with rectal cancer. *J Clin Oncol*. 2001;19:157-163.
6. Wong JH, Severino R, Honnebler MB, et al. Number of nodes examined and staging accuracy in colorectal carcinoma. *J Clin Oncol*. 1999;17:2896-2900.
7. Goldstein NS, Sanford W, Coffey M, et al. Lymph node recovery from colorectal resection specimens removed for adenocarcinoma: trends over time and a recommendation for a minimum number of lymph nodes to be recovered. *Am J Clin Pathol*. 1996;106:209-216.
8. Hammond ME, Fitzgibbons PL, Compton CC, et al, for the College of American Pathologists Cancer Committee and conference participants. College of American Pathologists conference XXXV: solid tumor prognostic factors: which, how and so what? Summary document and recommendations for implementation. *Arch Pathol Lab Med*. 2000;124:958-965.
9. Greene FL, Page DL, Fleming, et al (American Joint Commission on Cancer). *AJCC Cancer Staging Manual*. 6th ed. Philadelphia, PA: Lippincott-Raven; 2002:113-119.
10. Johnson PM, Malatjalian D, Porter GA. Adequacy of nodal harvest in colorectal cancer: a consecutive cohort study. *J Gastrointest Surg*. 2002;6:883-889.
11. Leibl S, Tsybrovskyy O, Denk H. How many lymph nodes are necessary to stage early and advanced adenocarcinoma of the sigmoid colon and upper rectum? *Virchows Arch*. 2003;443:133-138.
12. Scott KW, Grace RH, Gibbons P. Five-year follow-up study of the fat clearance technique in colorectal carcinoma. *Dis Colon Rectum*. 1994;37:126-128.
13. Scott KW, Grace RH. Detection of lymph node metastases in colorectal carcinoma before and after fat clearance. *Br J Surg*. 1989;76:1165-1167.
14. Brown HG, Luckasevic TM, Medich DS, et al. Efficacy of manual dissection of lymph nodes in colon cancer resections. *Mod Pathol*. 2004;17:402-406.
15. Haboubi NY, Clark P, Kaftan SM, et al. The importance of combining xylene clearance and immunohistochemistry in the accurate staging of colorectal carcinoma. *J R Soc Med*. 1992;85:386-388.
16. Newell KJ, Sawka BW, Rudrick BF, et al. An inexpensive, simple, and effective aid for the retrieval of lymph nodes from colorectal cancer resections. *Arch Pathol Lab Med*. 2001;125:642-645.
17. Sanchez W, Luna-Perez P, Alvarado I, et al. Modified clearing technique to identify lymph node metastases in post-irradiated surgical specimens from rectal adenocarcinoma. *Arch Med Res*. 1996;27:31-36.
18. Herrera L, Villarreal JR. Incidence of metastases from rectal adenocarcinoma in small lymph nodes detected by a clearing technique. *Dis Colon Rectum*. 1992;35:783-788.
19. Wyk QV, Hosie KB, Balsitis M. Histopathological detection of lymph node metastases from colorectal carcinoma. *J Clin Pathol*. 2000;53:685-687.
20. Liefers GJ, Cleton-Jansen AM, van de Velde CJ, et al. Micrometastases and survival in stage II colorectal cancer. *N Engl J Med*. 1998;339:223-228.
21. Merrie AE, van Rij AM, Dennett ER, et al. Prognostic significance of occult metastases in colon cancer. *Dis Colon Rectum*. 2003;46:221-231.
22. Herrera-Ornelas L, Justiniano J, Castillo N, et al. Metastases in small lymph nodes from colon cancer. *Arch Surg*. 1987;122:1253-1256.
23. Andreola S, Ermanno L, Filiberto B, et al. Manual dissection of adenocarcinoma of the lower third of the rectum specimens for detection of lymph node metastases smaller than 5 mm. *Cancer*. 1996;77:607-612.
24. Goldstein NS. Lymph node recoveries from 2427 pT3 colorectal resection specimens spanning 45 years: recommendations for a minimum number of recovered lymph nodes based on predictive probabilities. *Am J Surg Pathol*. 2002;26:179-189.
25. Wright FC, Law CHL, Last LD, et al. Barriers to optimal assessment of lymph nodes in colorectal cancer specimens. *Am J Clin Pathol*. 2004;121:663-670.