



Deposited via The University of Leeds.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/101977/>

Version: Accepted Version

Article:

Munkedal, DLE, Laurberg, S, Hagemann-Madsen, R et al. (2016) Significant individual variation between pathologists in the evaluation of colon cancer specimens after complete mesocolic excision. *Diseases of the Colon and Rectum*, 59 (10). pp. 953-961. ISSN: 0012-3706

<https://doi.org/10.1097/DCR.0000000000000671>

© The American Society of Colon & Rectal Surgeons, Inc. Unauthorized reproduction of this article is prohibited. This is an author produced version of a paper published in *Diseases of the Colon and Rectum*. Uploaded in accordance with the publisher's self-archiving policy. The final publication is available at Springer via [10.1097/DCR.0000000000000671](https://doi.org/10.1097/DCR.0000000000000671)

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Title:

Significant individual variation between pathologists in the evaluation of colon cancer specimens after complete mesocolic excision

Running head:

Variation in pathological evaluation

Authors:

Ditte Louise E. Munkedal ^a	Ditte.Munkedal@viborg.rm.dk	MD
Søren Laurberg ^a	soeren.laurberg@aarhus.rm.dk	Professor, DMSci
Rikke Hagemann-Madsen ^b	Rikke.Hjarno.Hagemann-Madsen@rsyd.dk	MD
Katrine J. Stribolt ^c	Katrine.Stribolt@rm.dk	MD
Søren R. P. Krag ^c	soerkrag@rm.dk	MD
Philip Quirke ^d	p.quirke@leeds.ac.uk	Professor, PhD
Nicholas P. West ^d	n.p.west@leeds.ac.uk	Ph.D

Affiliations:

a) Department of Surgery P, Aarhus University Hospital, 8000 Aarhus C, Denmark

b) Pathology Department, Vejle Regional Hospital, 7100 Vejle, Denmark

c) Pathology Department, Aarhus University Hospital, 8000 Aarhus C, Denmark

d) Pathology & Tumour Biology, Leeds Institute of Cancer and Pathology, University of Leeds, St. James's University Hospital, Leeds, LS9 7TF, UK

Corresponding author

MD Ditte Louise E. Munkedal, PhD-student
 Department of Surgery P
 Aarhus University Hospital
 Tage-Hansens Gade 2
 8000 Aarhus C, Denmark
 E-mail: Ditte.Munkedal@viborg.rm.dk
 Tel: +45 22 52 13 33

Disclaimers

Sources of support:

None

Word count:

Text: 2476

Abstract: 253

Author contribution:

Ditte Louise E. Munkedal ^a	Acquisition of data, analysis, interpretation of data, drafting the article and final approval
Søren Laurberg ^a	Contribution to design, acquisition of data, interpretation of data, revising the article and final approval
Rikke Hagemann-Madsen ^b	Acquisition of data, interpretation of data, revising the article and final approval
Katrine J. Stribolt ^c	Acquisition of data, interpretation of data, revising the article and final approval
Søren R. P. Krag ^c	Acquisition of data, interpretation of data, revising the article and final approval
Philip Quirke ^d	Contribution to design, interpretation of data, revising the article and final approval.
Nicholas P. West ^d	Contribution to design, acquisition of data, interpretation of data, revising the article and final approval

Category:

Colorectal/anal neoplasia

Key words:

Colon cancer, complete mesocolic excision, pathological evaluation, validation, surgical quality,

Abstract

Background: Following the introduction of complete mesocolic excision, a new pathological evaluation of the resected colon cancer specimen was introduced. This concept has quickly gained acceptance and is often used to compare surgical quality. The grading of colon cancer specimens is likely to depend on both surgical quality and the training of the pathologist.

Objective: To validate the principles of the pathological evaluation of colon cancer specimens.

Design: Exploratory study.

Settings: Aarhus, Denmark and Leeds, United Kingdom.

Patients: Colon cancers specimens.

Main outcome measures: The agreement of gradings between participants was of interest. Four specialist gastrointestinal pathologists and two abdominal surgeons evaluated two rounds of colon cancer specimens, each at two separate time points. Each round contained 50 specimens. Following the first round a protocol of detailed principles for the grading procedure was agreed upon. Results from an experienced pathologist were considered as the reference person.

Results: In the first round, the distribution of gradings between participants showed substantial variation. In the second round, the variation reduced. Intra-observer agreement was mostly fair-to-good whereas, inter-observer agreement was frequently poor. This did not significantly change from round one to round two.

Limitations: The small sample size of 100 specimens provided a very small number of specimens resected in the muscularis propria plane, which renders the evaluation of this group potentially unreliable. The evaluations were made on photos and not on fresh specimens.

Conclusion: This study demonstrates significant variation in the pathological evaluation of colon cancer specimens. It demonstrates that it cannot be used in clinical studies and care should be taken when comparing results between different hospitals.

Introduction

Colon cancer continues to have a high incidence in the western world, but despite an increasing focus on the disease, the mortality has not markedly decreased over the last 20 years. With the introduction of complete mesocolic excision (CME), colon cancer surgery has taken a step towards a more standardized and optimized resectional approach. The principles of CME are obtained from optimal rectal cancer surgery, where the introduction of total mesorectal excision (TME) has improved rectal cancer outcomes markedly^{1,2}. The concept of CME is now broadly accepted and has been implemented as the standard procedure in many centers. CME involves adequate removal of all tumor draining lymph nodes within an intact fascial and peritoneal lined package. This means dissection in the mesocolic plane when separating the colon and mesocolon from the retroperitoneum as described by Hohenberger et al.³, a central tie of the tumor feeding artery and sufficient cranio/caudal resection of the large bowel. The first part is generally accepted, whilst the two last components are still debated⁴. A developing body of evidence suggests that resection in the mesocolic plane is beneficial for patient survival^{5,6}.

Along with the introduction of CME, a new pathological evaluation of the resected colon cancer specimen was introduced⁵. The principle of grading specimens according to the quality of resection in the mesocolic plane is also derived from the evaluation of TME specimens⁷, and provides a better description of the surgical resection. A CME resection should result in a specimen with a smooth and intact surface encompassing all potential routes for tumor spread. The pathological evaluation is also a unique opportunity for feedback from pathologists to surgeons in order to optimize surgery, and surgical departments can use the evaluation to report

their results after implementation of CME^{4,8-11}. This pathological evaluation has likewise been implemented at many hospitals, and in Denmark has recently become mandatory. The grading of colon cancer specimens depends on both the surgical quality and training of the pathologist.

At Aarhus University Hospital (AUH) in Denmark, CME was implemented in 2008 along with the pathological evaluation of the specimen. Pathologists at AUH were trained by Professor Phil Quirke from the University of Leeds at a post-graduate course for multidisciplinary teams.

The aim of this study was to validate the concept of pathological evaluation of the plane of surgery for colon cancer resections. In a blinded study we analyzed inter-observer and intra-observer variability for surgeons and pathologists using randomly selected specimens from AUH. One of the developers of the pathological evaluation was used as the reference person.

Material and methods

Participants:

Four specialist gastrointestinal pathologists (NW, RHM, KS and SK) and two abdominal surgeons (DM and SL) were invited to participate in the study. None of the participants had performed the surgery but one of the pathologists from AUH may have seen some of the specimens during routine reporting years earlier. All pathologists at AUH were trained in the principles of grading colon cancer specimens, used it routinely and had participated in several international workshops on this subject. Both surgeons were very familiar with the principles of the pathological grading and had a short practical introduction by one of the pathologists. They did not routinely evaluate colon cancer specimens. The results from NW were considered as the reference.

Specimens and photographs for the study:

As it is not possible for multiple pathologists and surgeons to review the actual specimens without disrupting the diagnostic pathway, photographs were reviewed in line with previous studies of mesocolic grading. Two groups of 50 specimens were selected from a database covering 354 specimens from AUH resected between January 2008 and December 2011¹². We randomly selected specimens using www.randomizer.org adjusting for the location of the tumor within the colon. For all 100 specimens, whole fresh, whole formalin-fixed, and cross sectional slice photographs were obtained and available for review. All photographs were prospectively collected at the unit and retrospectively analyzed for the study. They were all high resolution and in color, demonstrating the front and back of the

specimen and a metric scale was visible in all photos. All patient identifiers were removed and a unique study code given.

Grading of specimens:

During both rounds, the participants had to determine whether each specimen was resected in the mesocolic plane, the intramesocolic plane or the muscularis propria plane.

The plane of surgery has previously been described but briefly consists of:

Mesocolic plane: A specimen with a fascial and peritoneal-lined surface with only very minor or no defects.

Intramesocolic plane: A specimen with a moderate amount of mesocolon and defects in the surface which are deeper than 5 mm, but do not reach the muscularis propria.

Muscularis propria: A specimen with at least one defect reaching the muscularis propria.

Rounds of grading:

For “Round 1”, all participants were asked to grade 50 specimens according to their current routine practice using the principles for the evaluations as described above.

The gradings were performed twice with a minimum of three months between each assessment so that specific specimens and their characteristics would be forgotten.

Due to the unexpected large variation in “Round 1”, we decided that all participants should meet in person to go through any discrepancies in the results from “Round 1”.

We agreed in more uniform principles the grading procedure (Appendix 1).

Afterwards, "Round 2" was conducted with a different group of 50 specimens. Again the gradings were performed twice with a minimum of three months apart, using the new agreed uniform principles.

All specimens for "Round 1" were graded between March 2013 and February 2014 and for "Round 2" between November 2014 and June 2015.

Statistics:

Weighted kappa statistics were used in order to compare the observed proportion of inter- and intra-observer agreement. Kappa values greater than 0.75 were taken as representing excellent agreement, values between 0.4 and 0.75 as fair to good agreement and values less than 0.4 were taken as poor agreement¹³. It was not possible to calculate P-values for the differences between Kappa values, but the signed rank test was used to test for symmetry between participants. P-values less than 0.05 were considered statistically significant. We also computed the proportion of perfect agreement between observers.

Results

Overall grading:

One specimen was excluded because one pathologist found it impossible to grade in “Round 1”, leaving 99 specimens for the final analysis.

According to the gold standard assessment, 67% of all specimens (both rounds) were resected in the mesocolic plane, 30% in the intramesocolic plane and 3% in the muscularis propria plane.

Distribution of gradings:

The individual distribution of the gradings for all rounds is shown in Figure 1.

In “Round 1” there was substantial variation between participants ranging from 63% of specimens resected in the mesocolic plane to 29% (Fig 1a-b). NW was very consistent in his grading but three out of six of the participants showed a significant change in the distribution of the gradings from assessment one to assessment two (Fig. 1b). In “Round 2” there was less variation in the gradings and there was no significant difference for any participant between assessment one and assessment two (Fig 1c-d).

Intra-observer variability of surgeons and pathologists:

In “Round 1” we found that NW had excellent agreement between the first and second assessment whereas the rest of the participants had fair to good agreement (Table 1a). The proportion of perfect agreement between the two evaluations was good across the group ranging from 71% to 90%. In “Round 2”, all comparisons of the first and second assessment represented a fair to good agreement. Figure 2 illustrates the complete intra-observer variability for NW for both rounds.

Inter-observer variability of surgeons and pathologists:

Table 2a represents the inter-observer variability compared to the reference for "Round 1". In more than half of the comparisons, the agreement was poor, however, the proportion of perfect agreement varied between 63% and 79% suggesting that at least around two thirds of the specimens were graded the same. In "Round 2", again more than half of the comparisons had poor agreement and the proportion of perfect agreement varied between 58% and 76%. The overall level of agreement is shown in Figure 3. In each evaluation, the agreement was around 70% for both the mesocolic plane and intramesocolic plane, whereas it was less than 50% for the muscularis propria plane.

Discussion

In this study we wanted to evaluate the intra- and inter-observer agreement between participants who grade colon cancer specimens as part of the clinical service on a regular basis. In “Round 1”, all specimens were evaluated according to current routine practice. Here we found substantial variation in the distribution of specimens classified in the mesocolic plane between surgeons and pathologists, and for some participants after repeating the evaluation. These results lead to a consensus meeting in order to develop uniform principles for the grading procedure, and for “Round 2” the variation in grading distribution was less pronounced. In both “Round 1” and “Round 2”, the agreement between participants showed substantial variation (Kappa values ranged from 0.14 to 0.58). The intra-observer variability was less pronounced than the inter-observer variability. A subgroup describing the difference between left and right sided specimens would have been interesting but was not possible due to the small sample size.

The major strength of this study is the strict methodology in how the participants evaluated each specimen and the use of one of the developers of mesocolic grading as the reference. The same specimen has been evaluated 6 times by pathologists and surgeons from two different centers which stress out the individual differences in the evaluation system. Limitations lie in the relatively small number of specimens evaluated. In particular, the small number of resections in the muscularis propria plane makes evaluation of this group potentially unreliable. As avoidance of the muscularis propria plane appears to be an important factor for patient survival, we recommend a larger study enriched for muscularis propria plane specimens to ensure consistency in the evaluation of suboptimal specimens. Evaluations in the

current study were performed on high resolution photographs of specimens however it is not clear whether the differences reported would be reproduced if the participants had prospective access to the fresh specimens and the ability to interrogate them from all angles. An evaluation based on fresh specimens would be optimal but was impossible to undertake in our study.

Following the proposal of CME as the optimal method of colon cancer resection, pathological grading of the specimen according to the plane of mesocolic dissection became a new way of evaluating the quality of surgery. The study from N. West et al. in 2008 became a key article and explained the principles for grading of CME specimens⁵. The surgical and pathological techniques were developed from TME surgery and evaluation, and the specimen is similarly described to be either in the mesocolic plane, intramesocolic plane or the muscularis propria plane. Through this evaluation, the standardized assessment of colon cancer specimens by pathologists makes it possible to compare the quality of surgery between surgeons and different institutions. To our knowledge, no other study has yet validated the histopathological evaluation of colon cancer specimens.

Colon cancer specimens show more varied anatomy due to the different regions of the colon when compared to rectal cancer specimens, so it might be expected that agreement in mesocolic grading would be worse than mesorectal grading. Two prior studies have shown variation in the pathological grading but only for TME specimens. A similar project has been undertaken for the evaluation of the mesorectal plane in rectal cancer resection¹⁴. Specimens assessed by local pathologists were re-evaluated by a review committee using digital photos. Like the present study, the concordance was low with a Kappa of 0.41 (95% CI, 0.30–0.52), even though they

re-evaluated a higher number of specimens. The conclusion is that there are a need for more objective and reproducible criteria in pathology.

The second study was a meta-analysis for the pathological evaluation of rectal cancer specimens, and demonstrated a substantial variation in plane of resection between centers ¹⁵, which could represent differences in the quality of the surgery rather than grading between pathologists.

This novel study illustrates two important issues regarding the grading of CME specimens.

First of all the evaluation of CME specimens is developed to reflect the surgical quality of the resection. The low Kappa values in the study reflect the fact that the principles of pathological specimen grading do not produce a truly objective score. Even the development of a uniform set of detailed instructions did not increase the Kappa values significantly. It is therefore difficult to agree on the true quality of a specimen. Despite this, the study shows a high level of intra-observer agreement for some of the participants, which implies that with experience, consistency in the evaluations will likely occur. Even with the level of subjectivity observed, specimen grading is felt to be an important component of surgical quality evaluation. It can be used to compare how surgeons in one unit change over time and provide individual feedback in postoperative multidisciplinary meetings by a review of specimen photographs. Similar subjective systems with significant intra-observer agreement are used in routine practice in colorectal cancer pathology to stratify patient care including tumor differentiation, tumor budding, regression grading after pre-operative treatment and vascular invasion.

The most important component of CME in order to improve patient survival is believed to be resection in the mesocolic plane. One study showed that resection in the muscularis propria plane was significantly associated with reduced overall survival following colon cancer resection⁵. Our study showed the worst level of agreement for the muscularis propria plane, however, there were very few specimens with major defects and therefore further investigation is needed.

The grading of mesocolic planes by pathologists does not currently have a direct consequence for the treatment of patients with colon cancer in Denmark. It is routinely implemented as a part of the Danish national guidelines¹⁶ in order to raise the surgical quality of colon cancer treatment. It is therefore mandatory for Danish pathologists to evaluate colon cancer specimens according to the resection plane and to record this in the pathological report.

The second issue of CME evaluation is the comparison of result between hospital units. The proportion of specimens resected in the mesocolic plane has become an accepted method of documenting the quality of colon cancer surgery between surgical departments in addition to other variables including lymph node yield, and has been presented in several articles detailing the introduction of CME. The uniformity of these evaluations for colon cancer has never been tested, and as this study shows, the gradings can vary substantially between individual pathologists. This degree of variation suggests that comparing results between units should be undertaken cautiously. The grading of the specimen appears to depend upon both the skill of the surgeons, but also the interpretation by the pathologist. In our study, we identified less variation in the proportion of specimens resected in the mesocolic

plane after the agreement of uniform grading principles (Appendix 1). This demonstrates that following detailed standardization of the guidelines, a comparison of specimens resected in the mesocolic plane between hospitals is a reasonable method of assessing variation in surgical quality.

The grading of colon cancer specimens is an important tool in measuring the quality of the surgery, but this is not an objective assessment and we require an internationally accepted uniform protocol in order to get more comparable results. This could be augmented by an online resource containing a series of good examples with a “test set”, in order to illustrate the evaluation of specimens and ensure consistency of results.

In conclusion, we have demonstrated significant variation in the grading of colon cancer specimens between pathologists. We suggest that the pathological grading of quality of colon cancer specimens requires further training and standardization prior to routine implementation in clinical practice. Therefore caution should be taken when comparing the results between different hospitals unless a detailed uniform protocol is followed.

Legends

Table 1a-b: Both tables show each participants agreement with themselves.

Numbers in parenthesis reflects a 95% confidence interval

Table 2a-b: The agreement between other participants and the reference person
(Nick West (NW))

Figure 1a-d: Individual distribution of grading in four rounds. P-values are derived from assessing the differences in the distribution of grading between the first and second assessment for each participant in the round.

Figure 2: Intra-observer variability for the reference person. All 99 specimens.

Figure 3: Inter-observer variability for each person compared to the reference person (50 specimens, four rounds, five comparisons). 990 evaluations as one specimen was excluded in Round 1.

References

1. Lemmens V, van Steenberghe L, Janssen-Heijnen M, Martijn H, Rutten H, Coebergh JW. Trends in colorectal cancer in the south of the Netherlands 1975-2007: rectal cancer survival levels with colon cancer survival. *Acta Oncol.* 2010;49:784-796.
2. Ostenfeld EB, Erichsen R, Iversen LH, Gandrup P, Norgaard M, Jacobsen J. Survival of patients with colon and rectal cancer in central and northern Denmark, 1998-2009. *Clin Epidemiol.* 2011;3 Suppl 1:27-34.
3. Hohenberger W, Weber K, Matzel K, Papadopoulos T, Merkel S. Standardized surgery for colonic cancer: complete mesocolic excision and central ligation--technical notes and outcome. *Colorectal Dis.* 2009;11:354-64; discussion 364-5.
4. Sondenaa K, Quirke P, Hohenberger W, et al. The rationale behind complete mesocolic excision (CME) and a central vascular ligation for colon cancer in open and laparoscopic surgery : Proceedings of a consensus conference. *Int J Colorectal Dis.* 2014;29:419-28
5. West NP, Morris EJ, Rotimi O, Cairns A, Finan PJ, Quirke P. Pathology grading of colon cancer surgical resection and its association with survival: a retrospective observational study. *Lancet Oncol.* 2008;9:857-865.
6. Bertelsen CA, Neuenschwander AU, Jansen JE, et al. Disease-free survival after complete mesocolic excision compared with conventional colon cancer surgery: a retrospective, population-based study. *Lancet Oncol.* 2014;16:161-8

7. Nagtegaal ID, van de Velde CJ, van der Worp E, et al. Macroscopic evaluation of rectal cancer resection specimen: clinical significance of the pathologist in quality control. *J Clin Oncol.* 2002;20:1729-1734.
8. Kobayashi H, West NP, Takahashi K, et al. Quality of Surgery for Stage III Colon Cancer: Comparison Between England, Germany, and Japan. *Ann Surg Oncol.* 2014;21 Suppl 3:S398-404
9. Gouvas N, Pechlivanides G, Zervakis N, Kafousi M, Xynos E. Complete Mesocolic Excision In Colon Cancer Surgery: A Comparison Between Open And Laparoscopic Approach Distal Right-Sided Colon Cancer Remains A Challenge For Laparoscopy. *Colorectal Dis.* 2012;14:1357-64
10. West NP, Kobayashi H, Takahashi K, et al. Understanding Optimal Colonic Cancer Surgery: Comparison of Japanese D3 Resection and European Complete Mesocolic Excision With Central Vascular Ligation. *J Clin Oncol.* 2012;30:1763-9
11. Bertelsen CA, Bols B, Ingeholm P, Jansen JE, Neuenschwander AU, Vilandt J. Can the quality of colonic surgery be improved by standardization of surgical technique with complete mesocolic excision? *Colorectal Dis.* 2011;13:1123-1129.
12. Munkedal DL, West NP, Iversen LH, Hagemann-Madsen R, Quirke P, Laurberg S. Implementation of complete mesocolic excision at a university hospital in Denmark: An audit of consecutive, prospectively collected colon cancer specimens. *Eur J Surg Oncol.* 2014;40:1494-501
13. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

14. Demetter P, Vandendael T, Sempoux C, et al. Need for objective and reproducible criteria in histopathological assessment of total mesorectal excision specimens: lessons from a national improvement project. *Colorectal Dis.* 2013;15:1351-1358.

15. Bosch SL, Nagtegaal ID. The Importance of the Pathologist's Role in Assessment of the Quality of the Mesorectum. *Curr Colorectal Cancer Rep.* 2012;8:90-98.

16. Danish Colorectal Cancer Group
Guidelines 2013
<http://dccg.dk/retningslinjer/indeks.html>. . January 2013.