Inter- and intraobserver variation in the histopathological evaluation of early oesophageal adenocarcinoma

Grotenhuis, B. A.; van Heijl, M.; ten Kate, F. J. W.; Biermann, K.; Offerhaus, G. J. A.; Wijnhoven, B. P. L.; van Berge Henegouwen, M. I.; Tilanus, H. W.; van Lanschot, J. J. B.

DOI
10.1136/jcp.2010.080721

Publication date
2010

Document Version
Final published version

Published in
Journal of clinical pathology

Citation for published version (APA):
Inter- and intraobserver variation in the histopathological evaluation of early oesophageal adenocarcinoma

Brechtje A Grotenhuis,1 Mark van Heijl,2 Fiebo J W ten Kate,3,4 Katharina Bierrmann,5 G Johan A Offerhaus,4 Bas P L Wijnhoven,1 Mark I van Berge Henegouwen,2 Hugo W Tilanus,1 J Jan B van Lanschot1

ABSTRACT

Aims According to the classification established by the Japanese Society for Oesophageal Disease, early oesophageal cancer can be subdivided into six successive layers of the mucosa or submucosa, which influences the treatment strategy and prognosis of the individual patient. However, the reproducibility of this classification in terms of inter- and intraobserver variability is unclear.

Methods Histological slides from 105 surgical resection specimens of patients who had undergone oesophagectomy for early oesophageal adenocarcinoma were reviewed independently by three gastrointestinal pathologists, and were classified according to the Japanese criteria (m1/m2/m3/sm1/sm2/sm3 tumours). Inter- and intraobserver variation was determined by k-statistics.

Results The interobserver reproducibility was good between pathologist 1 and 2 (k=0.61, 95% CI 0.55 to 0.67), and moderate between pathologist 1 and 3 (k=0.51, 95% CI 0.45 to 0.57) and between pathologist 2 and 3 (k=0.50, 95% CI 0.38 to 0.61). The intraobserver agreement as assessed by the expert pathologist was good (k=0.76), with a 95% CI that was interpreted as good to very good (0.67 to 0.85). Most agreement was achieved at the lower (m1) and upper site (sm2, sm3) of the spectrum, whereas the m2 tumours reflected the most discrepant stage. The majority of the observed discrepancy included the variation in one substage only.

Conclusions The reproducibility of the Japanese classification is good in terms of inter- and intraobserver variability when grading early oesophageal adenocarcinoma on surgical resection specimens. The present data confirm that dedicated gastrointestinal pathologists with broad experience are preferred when grading the resection specimens of patients with early oesophageal adenocarcinoma.

INTRODUCTION

Oesophageal adenocarcinoma is diagnosed as having increasing frequency at an early stage, due to the increased awareness of the clinical importance of Barrett’s oesophagus, more intensive surveillance programmes, and improved imaging techniques.1,2 According to the UICC (Union Internationale Contre le Cancer) TNM classification, early carcinoma is defined as an invasive tumour limited to the mucosa (T1a) or submucosa (T1b), irrespective of the lymph-nodal status.3 This dichotomy is also incorporated in the revised Vienna classification that has been established to resolve the discrepancies in nomenclature between Western and Japanese pathologists with regard to the grading of gastrointestinal epithelial neoplasia (table 1).4,5 In this classification, intramucosal carcinoma (together with high-grade dysplasia and carcinoma in situ set in category 4) is separated from submucosal invasion (category 5).4 Moreover, in 2001, the Japanese Society of Oesophageal Disease introduced its classification in which early cancers are subdivided into six successive layers of the mucosa (m1, m2, m3) or submucosa (sm1, sm2, sm3).6 However, the usefulness and reproducibility of this more recently introduced, extended classification have not yet been studied.

In general, a classification should be simple, a reflection of available therapeutic options and reproducible. Simplicity means that a classification should have as few categories as possible, but as many as required clinically. Reflection of available therapeutic options should include the possibilities of careful observation, local endoscopic treatment or surgery. It appears that subdivision in six, rather than two, categories is helpful in directing patients to the most optimal treatment strategy. Current treatment options for early oesophageal cancer vary from endoscopic mucosal resection (EMR) to surgical resection with extended or regional lymphadenectomy.7–10 The decision of whether to perform an endoscopic or a surgical resection is mainly dependent on patients’ predicted lymph-nodal status,11 which is related to the depth of infiltration of the primary tumour. Lymph-node metastases have not been reported in m1 and m2 mucosal adenocarcinomas (thereby allowing endoscopic treatment), whereas positive lymph nodes can be found in 0–12% of m3 patients.12–15 In submucosal adenocarcinomas, it appears that sm2–sm3 tumours are associated on average with a higher rate of lymph-node metastases than sm1 tumours12–15; hence, an oesophagectomy is considered the standard therapy.9,16–17 The optimal management of carcinomas with invasion into the muscularis mucosae (m3) or the most superficial layer of the submucosa (sm1) is still under debate. Finally, subdivision of early oesophageal cancer in six categories seems justified, as the treatment algorithm depends largely on the chance for lymphatic dissemination that is mainly related to the tumour infiltration depth.

Reproducibility of a classification includes the aspects of inter- and intraobserver variability. Thus
Table 1 Revised Vienna classification, established to resolve the discrepancies in nomenclature between Western and Japanese pathologists with regard to the grading of gastrointestinal epithelial neoplasia

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Negative for neoplasia/dysplasia</td>
<td>1</td>
</tr>
<tr>
<td>Indefinite for neoplasia/dysplasia</td>
<td>2</td>
</tr>
<tr>
<td>Mucosal low-grade neoplasia/dysplasia</td>
<td>3</td>
</tr>
<tr>
<td>Mucosal high-grade neoplasia</td>
<td></td>
</tr>
<tr>
<td>High-grade adenoma/dysplasia</td>
<td>4.1</td>
</tr>
<tr>
<td>Non-invasive carcinoma (carcinoma in situ)</td>
<td>4.2</td>
</tr>
<tr>
<td>Suspicious for invasive carcinoma</td>
<td>4.3</td>
</tr>
<tr>
<td>Intramucosal carcinoma</td>
<td>4.4</td>
</tr>
<tr>
<td>Submucosal invasive carcinoma</td>
<td>5</td>
</tr>
</tbody>
</table>

Figure 1 Depth of tumour invasion based on the criteria proposed by the Japanese Society for Oesophageal Disease: according to the Vienna classification, m1, m2 and m3 lesions represent category 4, whereas sm1, sm2 and sm3 tumours are classified as category 5. Reprinted with permission, all rights reserved © 2006 The American Journal of Surgery.

MATERIAL AND METHODS

Patient material

The outcome of 120 patients who underwent transhiatal oesophagectomy with regional lymphadenectomy for early adenocarcinoma of the distal oesophagus or gastro-oesophageal junction (pT1-tumours) between 1980 and 2002 in two university hospitals in The Netherlands (Erasmus Medical Centre, Rotterdam, and Academic Medical Centre, Amsterdam) has been described previously. Patients did not receive neoadjuvant chemo/radiotherapy. Archival material of the resection specimens was used for the present study. Tissue blocks and slides of the resection specimens of 13 patients were not available anymore, and in the resection specimens of another two patients the tumour was classified as a pt2 tumour (infiltration into the muscularis propria) by all three pathologists. Finally, resection specimens of 105 patients were used in the present study.

Histopathological assessment

Tumours were assigned pathological tumour-node-metastasis stages according to the UICC 2002 system. The depth of tumour invasion was measured and subclassified based on the criteria proposed by the Japanese Society for Oesophageal Disease (figure 1). High-grade dysplasia or carcinoma in situ without invasion through the basement membrane was classified as m1. Intramucosal carcinoma was defined as a tumour extending beyond the basement membrane into the lamina propria (m2). All carcinomas with invasion into the muscularis mucosae were classified as m3 tumours, irrespective of the presence of a double muscularis mucosae (ie, a superficial and a deep layer observed in Barrett’s mucosa can be classified as an m3 and m4 lesion, respectively). Carcinomas infiltrating beyond the (deeper) muscularis mucosae but limited to the upper third of the submucosa, were classified as sm1. Equally, carcinomas infiltrating the submucosa but limited to the middle or lower third of the submucosa were defined as sm2 or sm3, respectively.

To assess interobserver agreement in analysing early oesophageal cancer, histological slides of the primary tumour of 105 resection specimens were independently reviewed by three gastrointestinal pathologists. Pathologist 1 and pathologist 2 were highly experienced (more than 20 years’ experience) and had been closely working together in the same department. The third pathologist has a shorter work experience as a dedicated gastrointestinal pathologist (2 years), but in daily practice evaluates all pathology specimens with regard to oesophageal cancer in a high-volume centre. The pathologists were blinded for the identity of the patient and the initial diagnosis. The resection specimen of each individual patient was categorised based on the deepest infiltration of the primary tumour according to the six classes of early oesophageal cancer (m1, m2, m3, sm1, sm2, sm3). To assess intraobserver agreement, the most experienced pathologist who is considered an expert on oesophageal cancer rereviewed all histological slides 2 weeks after this first assessment.

Statistics

Inter- and intraobserver agreement were determined by k statistics. k statistics are widely used and accepted coefficients that provide a measure of observer agreement accounting for agreement other than that which occurs by chance alone. Coefficients <0.21, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80 and 0.81 to 1.00 represent a poor, fair, moderate, good and very good agreement, respectively. k-statistics with corresponding 95% CIs were calculated with regard to interobserver agreement (pathologist 1 versus pathologists 2 and 3, respectively, and pathologist 2 vs pathologist 3) and intraobserver agreement (as assessed by pathologist 1). Data analysis was carried out with SPSS version 15.0 (SPSS, Chicago, Illinois).

RESULTS

Clinicopathological characteristics of the present study population are shown in table 2. The infiltration depth of the primary tumour was initially classified as carcinoma in situ in 10 patients (9.5%), and as invasive carcinoma in 95 patients (90.5%). One m3 patient, five sm2 patients and 11 sm3 patients were diagnosed as having a N1 status.

The interobserver variation has been tested in three ways: pathologist 1 (expert on oesophageal cancer) versus pathologist 2, pathologist 1 versus pathologist 3 and pathologist 2 versus pathologist 3. The outcome with regard to this interobserver agreement other than that which occurs by chance alone.

Inter- and intraobserver agreement were determined by k statistics. k statistics are widely used and accepted coefficients that provide a measure of observer agreement accounting for agreement other than that which occurs by chance alone. Coefficients <0.21, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80 and 0.81 to 1.00 represent a poor, fair, moderate, good and very good agreement, respectively. k-statistics with corresponding 95% CIs were calculated with regard to interobserver agreement (pathologist 1 versus pathologists 2 and 3, respectively, and pathologist 2 vs pathologist 3) and intraobserver agreement (as assessed by pathologist 1). Data analysis was carried out with SPSS version 15.0 (SPSS, Chicago, Illinois).

RESULTS

Clinicopathological characteristics of the present study population are shown in table 2. The infiltration depth of the primary tumour was initially classified as carcinoma in situ in 10 patients (9.5%), and as invasive carcinoma in 95 patients (90.5%). One m3 patient, five sm2 patients and 11 sm3 patients were diagnosed as having a N1 status.

The interobserver variation has been tested in three ways: pathologist 1 (expert on oesophageal cancer) versus pathologist 2, pathologist 1 versus pathologist 3 and pathologist 2 versus pathologist 3. The outcome with regard to this interobserver agreement other than that which occurs by chance alone.

Inter- and intraobserver agreement were determined by k statistics. k statistics are widely used and accepted coefficients that provide a measure of observer agreement accounting for agreement other than that which occurs by chance alone. Coefficients <0.21, 0.21 to 0.40, 0.41 to 0.60, 0.61 to 0.80 and 0.81 to 1.00 represent a poor, fair, moderate, good and very good agreement, respectively. k-statistics with corresponding 95% CIs were calculated with regard to interobserver agreement (pathologist 1 versus pathologists 2 and 3, respectively, and pathologist 2 vs pathologist 3) and intraobserver agreement (as assessed by pathologist 1). Data analysis was carried out with SPSS version 15.0 (SPSS, Chicago, Illinois).
variation is shown in table 3. The interobserver reproducibility between pathologists 1 and 2 was good ($k=0.61$), with a 95% CI that can be interpreted as moderate to good (0.55 to 0.67). The interobserver agreement was graded as moderate between pathologists 1 and 3 ($k=0.51$, 95% CI 0.38 to 0.61) and between pathologists 2 and 3 ($k=0.50$, 95% CI 0.38 to 0.61). Furthermore, the most experienced pathologist rereviewed all histological slides to assess the intraobserver variation of this classification. The intraobserver agreement was good with a $k$ of 0.76, and a concomitant 95% CI that was interpreted as good to very good (0.67 to 0.85).

When analysing the discrepancies between the observers with regard to the six different substages as shown in table 4, it appeared that most agreement was achieved at the lower (m1) and upper site (sm2, sm3) of the spectrum. The m2 stage was the most discrepant stage (interobserver agreement 17% and 25% only), whereas in the m3 and sm1 stage, pathologist 3 showed less agreement. Furthermore, there was no underdiagnosis of submucosal carcinoma that requires undoubtedly surgical resection as only little discordance in the sm2 and sm3 substages was noted, and in 16 out of 17 N1 patients an sm2 or sm3 tumour was observed. It can be questioned if the discrimination in tumour infiltration depth will reflect the treatment options of endoscopic treatment (HGD, T1a) and surgery (T1b). When evaluating surgical resection specimens, the interobserver agreement was moderate when comparing HGD and intramucosal carcinoma, and good in case separation of intramucosal carcinoma from submucosal carcinoma was aimed for.

When preoperative biopsies were analysed, these agreements appeared to be only fair and poor, respectively. However, it has been suggested that interobserver agreement on EMR specimens is significantly higher than on biopsy specimens. Nevertheless, it may be a drawback of this study that interobserver variability has been evaluated from surgical resection specimens rather than from EMR specimens, as the latter will largely influence clinical decision-making.

To our knowledge, this is the first study that reports on the reproducibility of the Japanese classification in grading early oesophageal adenocarcinoma into six categories. In the present series, the m2 stage was the most discrepant stage (interobserver agreement 17% and 25% only), although it can be questioned if the interpretation of an m2 tumour as an m1 or m3 lesion will affect largely the clinical decision-making. In the m3 and sm1 stages, of which optimal treatment strategies are still under debate, pathologist 3 showed less agreement. Furthermore, there was no underdiagnosis of submucosal carcinoma that requires undoubtedly surgical resection as only little discordance in the sm2 and sm3 substages was noted, and in 16 out of 17 N1 patients an sm2 or sm3 tumour was observed. It can be questioned whether the differences in agreement between pathologist 1 and 2 versus the agreement between pathologist 3 and pathologists 1 and 2, respectively, reflect a difference in years of working experience, or whether the results are influenced by the fact that pathologists 1 and 2 had worked closely together in the same department for many years. Nevertheless, the present data confirm that dedicated gastrointestinal pathologists are preferred when grading the resection specimens of this subpopulation of patients with oesophageal cancer.

The Japanese classification for grading early oesophageal cancer plays an important role in the phase prior to the onset of treatment as well as the time period after the patient has undergone treatment. In the post-treatment phase, the final tumour stage as

### Table 2 Clinicopathological characteristics of 105 patients who underwent oesophagectomy for carcinoma in situ (m1) or early oesophageal adenocarcinoma (m2, m3, sm1, sm2 or sm3)

<table>
<thead>
<tr>
<th>Age*</th>
<th>66 years (30–82)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male 92 (87.6%) Female 13 (12.4%)</td>
</tr>
<tr>
<td>Tumour location</td>
<td>Distal oesophagus 93 (88.6%) Gastro-oesophageal junction 12 (11.4%)</td>
</tr>
<tr>
<td>Differentiation grade</td>
<td>Gx (undetermined) 7 (6.7%) G1 (good) 20 (19.0%) G2 (moderate) 59 (56.2%) G3 (poor) 19 (18.1%)</td>
</tr>
<tr>
<td>Barrett metaplasia</td>
<td>Yes 95 (90.5%) No 10 (9.6%)</td>
</tr>
<tr>
<td>Lymph-nodal involvement</td>
<td>pN0 88 (83.8%) pN1 17 (16.2%)</td>
</tr>
<tr>
<td>Radicallity of resection (R0)</td>
<td>105 (100%)</td>
</tr>
</tbody>
</table>

*Value presented as median (range in parentheses).

### Table 3 Inter- and intraobserver variation for the diagnosis of carcinoma in situ (m1) or early oesophageal adenocarcinoma (m2, m3, sm1, sm2 or sm3) as determined by the histopathological analysis of 105 resection specimens according to the Japanese classification

| Interobserver variation | Pathologist 1 versus pathologist 2 0.61 0.55 to 0.67 Good (moderate—good) Pathologist 1 versus pathologist 3 0.51 0.45 to 0.57 Moderate (moderate) Pathologist 2 versus pathologist 3 0.50 0.38 to 0.61 Moderate (fair—good) |  |
|-------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| Intraobserver variation | Pathologist 1 versus pathologist 1 0.76 0.67 to 0.85 Good (good—very good) |  |  |  |

*Interpretation of $k$-statistics, with interpretation of 95% CI in parentheses.

### Table 4 Inter- and intraobserver agreement as assessed per tumour stage, defined by the depth of tumour infiltration according to the Japanese classification

| Pathologist 1 | Interobserver Pathologist 2 Pathologist 3 Pathologist 1 |
|---------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|-------------------------------------------------|
| m1 (n=5) 4/5 (80%) 5/5 (100%) 5/5 (100%) |  |  |  |  |
| m2 (n=12) 2/12 (17%) 3/12 (25%) 11/12 (92%) |  |  |  |  |
| m3 (n=39) 29/39 (74%) 23/39 (59%) 28/39 (72%) |  |  |  |  |
| sm1 (n=18) 13/18 (72%) 8/18 (44%) 16/18 (89%) |  |  |  |  |
| sm2 (n=16) 12/16 (75%) 12/16 (75%) 13/16 (81%) |  |  |  |  |
| sm3 (n=15) 13/15 (87%) 13/15 (87%) 12/15 (80%) |  |  |  |  |

*Observations by pathologist 1 were considered the gold standard in this evaluation.

### DISCUSSION

The reproducibility of the classifications that aim for the discrimination between Barrett metaplasia, dysplasia and adenocarcinoma has been studied previously. Various groups have demonstrated that the use of such classifications (eg, Vienna classification) is still accompanied by considerable interobserver variability.

Furthermore, studies have been undertaken to assess the observer variation in the diagnosis of high-grade dysplasia, intramucosal adenocarcinoma (T1a) or submucosal adenocarcinoma (T1b). In general, this discrimination in tumour infiltration depth will reflect the treatment options of endoscopic treatment (HGD, T1a) and surgery (T1b). When evaluating surgical resection specimens, the interobserver agreement was moderate when comparing HGD and intramucosal carcinoma, and good in case separation of intramucosal carcinoma from submucosal carcinoma was aimed for. When preoperative biopsies were analysed, these agreements appeared to be only fair and poor, respectively. However, it has been suggested that interobserver agreement on EMR specimens is significantly higher than on biopsy specimens.

Nevertheless, it may be a drawback of this study that interobserver variability has been evaluated from surgical resection specimens rather than from EMR specimens, as the latter will largely influence clinical decision-making.

To our knowledge, this is the first study that reports on the reproducibility of the Japanese classification in grading early oesophageal adenocarcinoma into six categories. In the present series, the m2 stage was the most discrepant stage (interobserver agreement 17% and 25% only), although it can be questioned if the interpretation of an m2 tumour as an m1 or m3 lesion will affect largely the clinical decision-making. In the m3 and sm1 stages, of which optimal treatment strategies are still under debate, pathologist 3 showed less agreement. Furthermore, there was no underdiagnosis of submucosal carcinoma that requires undoubtedly surgical resection as only little discordance in the sm2 and sm3 substages was noted, and in 16 out of 17 N1 patients an sm2 or sm3 tumour was observed. It can be questioned whether the differences in agreement between pathologists 1 and 2 versus the agreement between pathologist 3 and pathologists 1 and 2, respectively, reflect a difference in years of working experience, or whether the results are influenced by the fact that pathologists 1 and 2 had worked closely together in the same department for many years. Nevertheless, the present data confirm that dedicated gastrointestinal pathologists are preferred when grading the resection specimens of this subpopulation of patients with oesophageal cancer.

The Japanese classification for grading early oesophageal cancer plays an important role in the phase prior to the onset of treatment as well as the time period after the patient has undergone treatment. In the post-treatment phase, the final tumour stage as
determined by histology of the resection specimen will define the patient’s prognosis. In the pretreatment phase, its role is even more important, as the decision whether to perform an endoscopic or a surgical resection in patients with early oesophageal cancer is mainly dependent on patients’ predicted lymph-nodal status. At present, all diagnostic modalities including endoscopic ultrasonography lack the ability to correctly stage the N-category with a very high accuracy (71–86% in literature). Nevertheless, on the basis of the T-stage as staged by the pathologist on an EMR specimen, one can predict the chance for lymphatic dissemination because of its relationship to the tumour infiltration depth as assessed by the Japanese classification. EMR is advocated as a staging modality to determine the T-stage in patients with early oesophageal cancer although the assessment of the exact depth of submucosal invasion remains difficult, as full-thickness submucosa is virtually absent in these specimens. Therefore, if the submucosal invasion exceeds 0.5 mm, one may consider the tumour infiltration beyond the sm1 layer. Other factors that may predict lymphatic involvement in early oesophageal cancer include a diameter greater than 3 cm, a poor differentiation grade and lymphangio-invasion. Although indications for endoscopic (m1, m2 tumours) and surgical treatment (sm2, sm3 tumours) are clear, for each individual with an m3 or sm1 adenocarcinoma it should be carefully considered whether a radical oesophagectomy may be beneficial. Therefore, individual treatment plans and a close interdisciplinary cooperation between surgeon, gastroenterologist and pathologist are mandatory in this patient population.

In conclusion, the reproducibility of the Japanese classification system is good in terms of inter- and intraobserver variability when grading early oesophageal adenocarcinoma on surgical resection specimens. The present data confirm that dedicated gastrointestinal pathologists with broad experience are preferred when grading the resection specimens of patients with early oesophageal adenocarcinoma.

Acknowledgements We would like to thank SE Hoeks (Erasmus MC, Department of Anaesthesiology) for her statistical advice.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES
Inter- and intraobserver variation in the histopathological evaluation of early oesophageal adenocarcinoma

Brechtje A Grotenhuis, Mark van Heijl, Fiebo J W ten Kate, et al.

*J Clin Pathol* 2010 63: 978-981 originally published online September 23, 2010
doi: 10.1136/jcp.2010.080721