

Interobserver variation in the histopathological scoring of *Helicobacter pylori* related gastritis

Xiao-Yu Chen, Rene W M van der Hulst, Marco J Bruno, Arie van der Ende, Shu-Dong Xiao, Guido N J Tytgat, Fiebo J W Ten Kate

Abstract

Aim—To test the reproducibility between two histopathologists of features of *Helicobacter pylori* gastritis, using the updated Sydney classification.

Methods—290 dyspeptic Dutch patients with biopsy proven *H pylori* infection were enrolled in the study. Gastric antral mucosal biopsy specimens were analysed before and after *H pylori* eradication treatment. The biopsies were scored semi-quantitatively by two histopathologists, according to the updated Sydney classification system. Variables analysed included the density of *H pylori* infection, the degree of chronic inflammation, inflammatory activity, atrophy, intestinal metaplasia, and surface epithelial damage. Before grading biopsy specimens, both pathologists reached a consensus on the scoring of gastritis through interactive sessions using a multihanded microscope. Subsequently all biopsy specimens were graded. Interobserver variability was also analysed using weighted κ scores.

Results—For interobserver agreement on scoring the various gastritis features a high degree of reproducibility was reached overall. Agreement on grading of atrophy was the lowest; however, moderate to good reproducibility was achieved, with weighted κ values of 0.49 in the pretreatment biopsies and 0.52 in the post-treatment biopsies. Disagreement was most common in biopsy specimens with lesser degrees of atrophy. A high degree of agreement was obtained for intestinal metaplasia, with weighted κ values of 0.72 in the pretreatment biopsies and 0.73 in the post-treatment biopsies. The best agreement was reached in the assessment of the density of *H pylori* both before and after *H pylori* eradication treatment, with excellent weighted κ values of 0.76 and 0.95, respectively. The grade of reproducibility of inflammatory activity, superficial epithelial damage, and chronic inflammation was high, with weighted κ values varying from 0.60 to 0.76 and 0.62 to 0.83 before and after eradication, respectively.

Conclusions—Reproducibility of grading *H pylori* related gastritis is high using the updated Sydney system. Despite the novel criteria for scoring atrophy, there was imperfect agreement on this feature between two independent histopathologists.

(*J Clin Pathol* 1999;52:612-615)

Keywords: *H pylori* gastritis; interobserver variation; Sydney classification

Since the discovery of *Helicobacter pylori* in the gastric mucosa by Warren and Marshall in 1983, several studies have shown that this organism is strongly associated with chronic active gastritis as well as gastric adenocarcinoma and MALToma.¹⁻³ However, the morphological criteria for classification and grading of chronic gastritis remain obscure and poorly standardised, especially with respect to atrophy. In an attempt to solve this problem, the Sydney classification of gastritis was devised in 1990,⁴ and it was later updated in 1994.^{5,6} A new visual analogue scale for grading morphological variables and a set of guidelines for its application were subsequently designed.⁷ Since then, several studies on interobserver variation on the assessment of *H pylori* gastritis have been reported⁸⁻¹⁰ and these are summarised in table 1. One of those studies revealed excellent interobserver agreement for density of *H pylori*, moderate agreement for activity of gastritis, and poor agreement for the degree of atrophy.⁹

Gastric atrophy is defined as the loss of gastric glands, and this occurs in distinct patterns. Loss of gastric glands may follow a multifocal distribution, usually accompanied by intestinal metaplasia. Glandular atrophy is relatively easy to detect when it is severe. However, for lesser degrees of atrophy, particularly in the antrum, actual loss of glands and apparent loss of glands as a result of separation by inflammatory cells are difficult to distinguish.^{6,11,12} Interpretation is especially difficult when tissue sampling is not adequate or if biopsies are not well oriented. Recently, it has been shown in several studies that even experienced gastrointestinal pathologists have poor interobserver agreement over the assessment of this feature of *H pylori* gastritis.^{6,8,9}

Most of these studies have been done in Europe. In China, the Sydney system has only been accepted in some areas; elsewhere reluctance prevails, partly because of the less than optimal grading criteria in the updated Sydney system. In the current study, we tested the value and interobserver agreement of the updated Sydney system for scoring *H pylori* gastritis when used by a Chinese and a Dutch pathologist. We compared assessment of the density of *H pylori*, the degree of inflammatory activity, chronic inflammation, atrophy, intestinal metaplasia, and in addition surface epithelial damage, especially with regard to interobserver variation.

Department of Pathology, Academic Medical Centre, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands
F J W Ten Kate

Department of Gastroenterology, Academic Medical Centre, University of Amsterdam
R W M van der Hulst
M J Bruno
G N J Tytgat

Department of Medical Microbiology, Academic Medical Centre, University of Amsterdam
A van der Ende

Shanghai Institute of Digestive Disease, Shanghai Second Medical University, Shanghai, People's Republic of China
X-Y Chen
S-D Xiao

Correspondence to:
Dr Ten Kate.
email:
f.j.tenkate@amc.uva.nl

Accepted for publication
1 April 1999

Table 1 κ Values for scoring different gastritis features in antral biopsies from published reports

| Report [ref No] | Criteria | n | κ Value | | | | | | |
|---------------------------------------|-------------------------------|-----|----------------|-----------|-------------|------|-----------|-----------|------|
| | | | Hp density | Activity | Chronic inf | SED | Atrophy | IM | |
| Andrew (UK) ⁸ | The Sydney system (0–3 scale) | 69 | 0.74 | 0.69 | 0.58 | – | 0.51 | 0.73 | |
| | Present or absent | 99 | 0.89–0.93 | 0.80–0.84 | – | – | 0.08–0.34 | – | |
| El-Zimaity (USA) ⁹ | 0–3 scale (the Sydney system) | | 0.68–0.77 | 0.55–0.61 | – | – | 0.19–0.29 | – | |
| | 0–5 scale* | | 0.58–0.67 | 0.42–0.48 | – | – | 0.06–0.16 | – | |
| Fiocca (Italy) ¹⁰ | The Sydney /Houston system | 200 | 0.62–0.86 | 0.58–0.77 | 0.49–0.82 | – | 0.42–0.59 | 0.75–0.92 | |
| Present study (weighted κ) | 0–6 scale** | 290 | Pre- | 0.76 | 0.76 | 0.60 | 0.69 | 0.49 | 0.72 |
| | | | Post- | 0.95 | 0.83 | 0.62 | 0.73 | 0.52 | 0.73 |

*A visual analogue scale on 0–5 point according to the Sydney system.

**A detailed histological classification (0–6 scale) according to the updated Sydney system.

Hp, *Helicobacter pylori*; IM, intestinal metaplasia; inf, inflammation; SED, superficial epithelial damage.

Methods

Two hundred and ninety consecutive dyspeptic Dutch patients with proven *H pylori* infection were enrolled in a long term follow up study conducted in the Academic Medical Centre, University of Amsterdam.¹³

Two to five gastric antral biopsy specimens from the lesser curvature and the greater curvature, within 2 cm of the fundal-pyloric border, were taken before and after *H pylori* eradication treatment. Overall 697 gastric antral biopsies were examined independently by two experienced pathologists, one from The Netherlands (FTK, observer 1) and one from China (X-YC, observer 2).

Before starting the study, the two pathologists reached consensus about the methods of grading the features of gastritis through interactive sessions at a multiheaded microscope.

HISTOLOGICAL ASSESSMENT OF GASTRITIS

Gastric biopsies from the antrum were paraffin embedded, sectioned at 4 μ m and stained with haematoxylin and eosin. These were scored semiquantitatively according to the updated Sydney classification.¹¹ The following histological features were examined on each slide: density of *H pylori*, inflammatory activity, chronic inflammation, atrophy, intestinal metaplasia, and surface epithelial damage. The pathologists were blinded to any clinical information.

The updated Sydney system has a scale of 0–3 for scoring the features of chronic gastritis.¹¹ In order to improve assessment of minor degrees of alteration, a detailed histopathological classification was used, which also provided numerical data for statistical analysis. Each category (mild, moderate, and severe) was divided into two subcategories, resulting in a score on a scale of 0–6 (none, 0; mild, 1–2; moderate, 3–4; severe, 5–6), corresponding to the updated Sydney system as previously described.¹⁰

Density of *H pylori* colonisation

The density of *H pylori* colonisation was graded as follows: 0, none; 1, *H pylori* found only in one place after a careful search; 2, only a few *H pylori* found; 3, scattered *H pylori* found in separate areas/foci; 4, numerous *H pylori* in separate areas/foci; 5, nearly complete gastric surface covered

by a layer of *H pylori*; 6, continuous gastric surface coverage by a thick layer of *H pylori*.

Degree of inflammatory activity

The degree of inflammatory activity was scored according to the density of neutrophils in the gastric mucosa: 0, none; 1, only one crypt involved per biopsy; 2, two crypts involved per biopsy; 3, many crypts (up to 25%) involved per biopsy; 4, 25–50% of crypts involved per biopsy; 5, more than 50% of crypts involved per biopsy; 6, all crypts involved.

Superficial epithelial damage

Superficial epithelial damage was scored as follows: 0, none; 1, slight; 2, mild degeneration in the top of the epithelial cells; 3, moderate degeneration with disorientation of the epithelial lining; 4, indistinct cell borders at the surface of the epithelium; 5, flattened epithelial cells with severe degeneration and enlarged nuclei; 6, flattened to erosive epithelium of the entire surface.

Degree of chronic inflammatory infiltrate

The degree of chronic inflammatory infiltrate in the gastric mucosa (lymphocytes, plasma cells) was scored as follows: 0, none; 1, scattered chronic inflammatory cells, less than 10 in each high power field; 2, scattered chronic inflammatory cells, > 10 cells /high power field; 3, some areas with dense chronic inflammatory cells; 4, diffuse infiltration with dense chronic inflammatory cells; 5, nearly the whole mucosa contains dense chronic inflammatory cells which separate the gastric glands; 6, entire mucosa contains a dense chronic inflammatory cell infiltrate.

Degree of intestinal metaplasia

The degree of intestinal metaplasia was graded according to the amount of glandular tissue replaced by intestinal-type epithelium: 0, none; 1, only one focus (one crypt) replaced by intestinal-type epithelium; 2, one focal area (1–4 crypts) in one of two biopsies; 3, two separate foci; 4, multiple foci in one or both biopsies; 5, more than 50% gastric epithelium diffusely replaced by intestinal metaplasia; 6, only a few small area of gastric epithelium are not replaced by intestinal metaplasia.

Table 2 The weighted κ values for interobserver agreement on features of histopathological gastritis using a 0–6 scale according to the updated Sydney classification

| Weighted κ value | Histopathological features | | | | | |
|-------------------------|----------------------------|-----------------------|----------------------|---------|------|------|
| | <i>H pylori</i> density | Inflammatory activity | Chronic inflammation | Atrophy | IM | SED |
| Pre- | 0.76 | 0.76 | 0.60 | 0.49 | 0.72 | 0.69 |
| Post- | 0.95 | 0.83 | 0.62 | 0.52 | 0.73 | 0.73 |

IM, intestinal metaplasia; SED, superficial epithelial damage.

Table 3 The grading of atrophy by two histopathologists on a 0–6 scale according to the updated Sydney classification

| | Observer 1 | | | | | | | Total (n) |
|------------|------------|----|----|----|----|---|---|-----------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | |
| Observer 2 | | | | | | | | |
| 0 | 58 | 49 | 44 | 11 | 8 | | | 167 |
| 1 | 3 | 5 | 11 | 2 | | | | 23 |
| 2 | 1 | 2 | 18 | 10 | 9 | | | 41 |
| 3 | | | 9 | 7 | 10 | 1 | 2 | 29 |
| 4 | | | | | 10 | 1 | 3 | 14 |
| 5 | | | | | 8 | 4 | 2 | 14 |
| 6 | | | | | | | 2 | 2 |
| Total (n) | 62 | 56 | 82 | 30 | 45 | 6 | 9 | 290 |

Table 4 Consensus diagnosis for the eight cases with major discrepancy in atrophy scores

| Case | Observer 1 | Observer 2 | Final consensus |
|------|------------|------------|-----------------|
| 1 | 0 | 4 | 2 |
| 2 | 0 | 4 | 3 |
| 3 | 0 | 4 | 1 |
| 4 | 0 | 4 | 1 |
| 5 | 0 | 4 | 3 |
| 6 | 0 | 4 | 2 |
| 7 | 0 | 4 | 0 |
| 8 | 0 | 4 | 1 |

Atrophy

Atrophy was defined as the loss of specialised gastric glandular tissue, partly characterised by replacement by intestinal metaplasia. The presence of intestinal metaplasia can be used as an indicator of the severity and extent of atrophy in antrum, but atrophy might occur without accompanying intestinal metaplasia (areas with lymphoid follicles or dense infiltrate, or both, should not be graded): 0, none; 1, foci where a few gastric glands are lost or replaced by intestinal-type epithelium; 2, small areas in which gastric glands have disappeared or been replaced by intestinal-type epithelium; 3, up to 25% gastric glands lost or replaced by intestinal-type epithelium; 4, 25–50% of gastric glands lost or replaced by intestinal-type epithelium; 5, more than 50% of gastric glands lost or replaced by intestinal-type epithelium; 6, only a few small areas of gastric glands remaining.

Table 5 The grading of *H pylori* density by two histopathologists on a 0–6 scale according to the updated Sydney classification

| | Observer 1 | | | | | | | Total (n) |
|------------|------------|---|----|----|----|----|----|-----------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | |
| Observer 2 | | | | | | | | |
| 0 | 163 | 3 | 1 | | | | | 167 |
| 1 | 4 | 4 | 4 | | | | | 12 |
| 2 | | 1 | 12 | 1 | | | | 17 |
| 3 | | | 12 | 7 | 18 | 1 | | 38 |
| 4 | | | | 4 | 8 | 17 | 4 | 33 |
| 5 | | | | 1 | | 5 | 12 | 18 |
| 6 | | | | | | | 6 | 6 |
| Total (n) | 167 | 8 | 29 | 16 | 26 | 23 | 22 | 290 |

STATISTICAL ANALYSES

We used the weighted κ statistic. This statistic, as opposed to the unweight κ statistic, accounts for the fact that if two observers differ by more than one step, their disagreement should be given more weight than if they differ by only one step. Weighted κ values greater than 0.70 are considered to represent excellent reproducibility, between 0.5 and 0.7 to represent good reproducibility, between 0.3 and 0.5 to represent moderate to good reproducibility, and less than 0.3 to represent poor reproducibility.

Results

Interobserver agreement on the various gastritis features and weighted κ values is shown in table 2. The overall interobserver agreement on the scoring of the various gastritis features was on average high. In general, the level of interobserver agreement was better in post-treatment biopsies than in pre-treatment biopsies.

Overall agreement for the grade of atrophy was lower than for the other gastritis features, but moderate to good reproducibility was still achieved, with κ values of 0.49 in the pretreatment biopsies and 0.52 in the post-treatment biopsies. Complete agreement on the grading of atrophy on 0–6 scales in our study was reached in 36% of the cases (104/290); one step disagreement occurred in 29% (85/290); two step disagreements were found in 28% (81/290), and disagreement of more than two steps was seen in 7% (21/290). When grading atrophy, observer 2 consistently scored one step higher than observer 1 in cases of disagreement. Most disagreements in the grading of atrophy occurred in biopsy specimens where the degree of atrophy was low overall (table 3). The cases in which interobserver discrepancies for atrophy were 0–4 were re-examined and a final consensus was reached (table 4).

A high degree of agreement was obtained with respect to intestinal metaplasia, with κ values of 0.72 in the pretreatment biopsies and 0.73 in the post-treatment biopsies.

The best agreement was obtained in the assessment of the density of *H pylori* before and after *H pylori* eradication treatment, with κ values of 0.76 and 0.95 respectively (table 5).

Agreement about neutrophil infiltration using semiquantitative scoring was also excellent, with κ values of 0.76 in the pretreatment biopsies and 0.83 in the post-treatment biopsies. The grading of superficial epithelial damage also reached good to excellent reproducibility, with weighted κ values of 0.69 and 0.73 for pre- and post-treatment biopsies, respectively. However, the κ value for assessment of the degree of chronic inflammation using semiquantitative scoring was lower than for acute inflammation and for surface epithelial damage, though good reproducibility was still achieved.

Discussion

On testing interobserver variation between a Chinese and a Dutch pathologist, we found almost perfect agreement over the evaluation of

the density of *H pylori* in post-treatment gastric mucosa ($\kappa = 0.95$), and excellent agreement in pretreatment gastric mucosa ($\kappa = 0.76$) was achieved. This is consistent with the study by El-Zimaity *et al.*,⁹ who also found excellent interobserver agreement for the presence or absence of *H pylori* ($\kappa = 0.90$) (table 1). The κ value for the semiquantitative assessment of *H pylori* in their study was lower, with a value of 0.60 on the six point scale and 0.73 on the 0–3 scale. In our study, post-treatment κ values were probably higher because approximately 50% of patients were *H pylori* negative after eradication treatment.

The degree of neutrophil infiltration, superficial epithelial damage, and chronic inflammation by semiquantitative scoring reached a good to excellent reproducibility, with κ values ranging from 0.76–0.60 in pretreatment biopsies to 0.83–0.62 in post-treatment samples. However, κ values for assessment of the degree of chronic inflammation were lower than those for the acute gastritis component and for surface epithelial damage. This is in contrast with the study by El-Zimaity *et al.*,⁹ who found moderate interobserver agreement for the presence or absence of neutrophil infiltration in the gastric mucosa ($\kappa = 0.80$). Also the κ value for the semiquantitative assessment of neutrophils was lower on the six point scale ($\kappa = 0.43$) than on the 0–3 scale ($\kappa = 0.54$).

Excellent reproducibility was reached regarding intestinal metaplasia, with κ values of 0.72 in the pretreatment biopsies and 0.73 in the post-treatment biopsies in our study. Our findings are in accordance with those of Fiocca *et al.*,¹⁰ who reported κ values ranging from 0.75 to 0.92 in the antral biopsies among three pathologists using the Sydney/updated Sydney criteria (table 1). Using the Sydney system, Andrew *et al.* also found a κ value for intestinal metaplasia of 0.73 in antral biopsies among three pathologists (table 1).⁸

The overall agreement for the grade of atrophy was the lowest among the gastritis features, but moderate to good reproducibility was still achieved, with κ values of 0.49 in the pretreatment biopsies and 0.52 in the post-treatment biopsies. The highest level of disagreement occurred when scoring biopsy specimens with a low level of atrophy overall. The main causes of disagreement were marked inflammatory infiltration, with huge lymphoid follicles occupying a large portion of the gastric mucosa, and small biopsy samples, in which technical disturbances interfered with proper interpretation. Observer 2 consistently scored one step higher than observer 1 in the cases of disagreement. Therefore consensus agreement and rigorous interobserver studies on the level of agreement are needed before starting an interobserver study. As in our study, others have also shown the lowest agreement for the scoring of atrophy (table 1), with κ values varying from 0.42 in the study by Fiocca *et al.* to 0.51 in the study by Andrew *et al.*^{8 10} El-Zimaity *et al.* also found the poorest agreement for atrophy, with κ value ranged from 0.08 to 0.29, which depended on the site of the biopsy.⁹ Disagreement over the scoring of atrophy

relates to differences in interpreting actual loss of glands and apparent loss of glands, as a result of the separation of the glandular structures by inflammatory infiltrate.^{6 11 12} If the sampling is not adequate or if the biopsies are not well oriented, adequate assessment is even more difficult.

The main pitfall in grading gastric atrophy is related to its precise definition. Gastric atrophy is defined as a loss of gastric glands, but lesser degrees of atrophy, particularly in the antrum, can be difficult to assess because clear criteria are lacking. The antral mucosa is composed of pits and glands and their branchings. The antral glands are shorter and less closely packed but more tortuous and branched than the glands in the corpus. Pits lined with surface mucous cells occupy much of the gland. The glands are coiled from the base of the pits to the muscularis mucosae. Normally, in well oriented biopsies, there are about three coils between the base of the pit and the muscularis mucosae.

In our study, four criteria may be helpful in judging the presence and the degree of atrophy in antrum: (1) the presence of intestinal metaplasia; (2) the number of coils; (3) disturbance of pit numbers in relation to those of glandular structure; (4) the presence of fibrosis.

In conclusion, the updated Sydney system for scoring *H pylori* gastritis is useful and reproducible, but there needs to be improvement in the criteria for grading atrophy.

This work was supported in part by a grant from the Dutch Ministry of Education and Science and the Royal Dutch Academic of Science (KNAW). We thank to Professor Jaap Simons for correcting the English language and Carin B Heringa (Yamanouchi Europe) for collecting the clinical data from China.

- 1 Van der Hulst RWM, Van der Ende A, Dekker FW, *et al.* Effect of *Helicobacter pylori* eradication on gastritis in relation to cagA: a prospective 1-year follow-up study. *Gastroenterology* 1997;113:25–30.
- 2 Correa P. *Helicobacter pylori* and gastric carcinogenesis. *Am J Surg Pathol* 1995;19(suppl):s37–43.
- 3 Bayerdorffer E, Neubauer A, Rudolph B, *et al.* Regression of primary gastric lymphoma of MALT after cure of *Helicobacter pylori* infection. *Lancet* 1995;345:1591–4.
- 4 Misiewicz JJ, Tytgat GNJ, Goodwin CS. The Sydney system: a new classification of gastritis. *J Hepatol Gastroenterol* 1991;6:209–22.
- 5 Genta RM, Dixon MF. The Sydney system revisited: the Houston international gastritis workshop. *Am J Gastroenterol* 1995;90:1039–41.
- 6 Genta RM. *Helicobacter pylori*, inflammation, mucosal damage, and apoptosis: pathogenesis and definition of gastric atrophy. *Gastroenterology* 1997;113:s51–5.
- 7 Vet HCW, Koudstaal J, Kwee WS, *et al.* Efforts to improve interobserver agreement in histopathological grading. *J Clin Epidemiol* 1995;48:869–73.
- 8 Andrew A, Wyatt JJ, Dixon MF. Observer variation in the assessment of chronic gastritis according to the Sydney system. *Histopathology* 1994;25:317–22.
- 9 El-Zimaity HMT, Graham DY, Al-assi MT, *et al.* Interobserver variation in the histopathological assessment of *Helicobacter pylori* gastritis. *Hum Pathol* 1996;27:35–41.
- 10 Fiocca R, Villani L, Cornaggia M, *et al.* Interobserver variation in the assessment of *H pylori* gastritis [abstract]. *Gut* 1996;39(suppl 2):A104–5.
- 11 Dixon MF, Genta RM, Yardley JH, *et al.* Classification and grading of gastritis, the updated Sydney system. *Am J Surg Pathol* 1996;20:1161–81.
- 12 Genta RM. Recognizing atrophy: another step toward a classification of gastritis. *Am J Surg Pathol* 1996;20(suppl 1):s23–30.
- 13 RWM van der Hulst, FJW ten Kate, EAJ Rauws, *et al.* The relation of CagA and the long-term sequelae of gastritis after successful cure of *H pylori* infection—a long-term follow-up study [abstract]. *Gastroenterology* 1998;114:A318.



Interobserver variation in the histopathological scoring of *Helicobacter pylori* related gastritis.

X Y Chen, R W van der Hulst, M J Bruno, et al.

J Clin Pathol 1999 52: 612-615

doi: [10.1136/jcp.52.8.612](https://doi.org/10.1136/jcp.52.8.612)

Updated information and services can be found at:

<http://jcp.bmj.com/content/52/8/612>

References

These include:

Article cited in:

<http://jcp.bmj.com/content/52/8/612#related-urls>

Email alerting service

Receive free email alerts when new articles cite this article. Sign up in the box at the top right corner of the online article.

Notes

To request permissions go to:

<http://group.bmj.com/group/rights-licensing/permissions>

To order reprints go to:

<http://journals.bmj.com/cgi/reprintform>

To subscribe to BMJ go to:

<http://group.bmj.com/subscribe/>