Gut 1994; **35:** 1030–1032

Interobserver variation in assessment of gastroduodenal lesions associated with non-steroidal anti-inflammatory drugs

N Hudson, S Everitt, C J Hawkey

Abstract

Video endoscopic images were used to investigate whether gastroenterologists could agree on the definition of lesions within the stomach seen at endoscopy, with particular reference to those seen in patients taking non-steroidal anti-inflammatory drugs. Seven experienced endoscopists, unaware of the patients' clinical history or drug consumption, recorded their classification for 93 randomised video images of gastric lesions. There was complete agreement in the diagnosis of ulceration for nine images from patients who were not taking non-steroidal anti-inflammatory drugs; eight of nine were classified as deep ulcers, with 86% agreement for this subclassification. By contrast, the overall agreement for lesions in patients taking non-steroidal anti-inflammatory drugs was only 55%. Only nine of 44 ulcers were subclassified as deep, and there was considerable cross classification of non-haemorrhagic erosions and ulcers. In conclusion, ulcers that occur in patients taking non-steroidal anti-inflammatory drugs differ from those in patients who are not taking these drugs in that they are often more superficial and difficult to distinguish from erosions. The prognostic importance of these lesions is, therefore, uncertain.

(*Gut* 1994; **35:** 1030–1032)

Endoscopic studies in patients taking aspirin or non-steroidal, anti-inflammatory drugs (NSAIDs) suggest a point prevalence of gastric or duodenal ulceration in excess of 20%.[1–8] This contrasts with the much lower rate of ulceration infered from controlled epidemiological studies of clinically significant end points such as haematemesis and melaena.[9 10] Such endoscopic studies often do not include a definition of ulceration or may use a limited one based upon an assertion that the lesion has 'depth', and is greater than 3 or 5 mm in diameter. These definitions, however, have not been subject to validation, for example by interobserver correlation and by comparison with ulcers found in patients who are not taking NSAIDs. This study aimed to investigate interobserver variations between experienced endoscopists in the assessment of gastric lesions in patients, with particular regard to those lesions associated with NSAID use.

Department of Therapeutics, University Hospital, Nottingham
N Hudson
S Everitt
C J Hawkey

Correspondence to:
Professor C J Hawkey, Department of Therapeutics, University Hospital, Nottingham NG7 2UH.

Accepted for publication 16 November 1993

Methods

Endoscopic video recordings of gastric lesions were recorded with an Olympus PV10 endoscope and CV-1 video processor. All erosive lesions encountered by one of the authors (NH) over a period of seven months were recorded. These were assessed without knowledge of drug use by two of the investigators and those not regarded as adequately clear were discarded. This left 93 consecutive video images of ulcers or erosions – 84 from patients participating in a study of the gastroduodenal effects of NSAIDs and nine from patients who were not taking NSAIDs and were being investigated concurrently for dyspepsia on the same endoscopy list. Informed consent was obtained before the endoscopy and video recordings. Continual recordings of appropriate lesions were made in a standardised manner for a period of between 20 and 30 seconds from several different angles and distances.

An initial classification of each lesion was made by the endoscopist (NH) and recorded. Lesions were classified as follows:

Category (0) *Erythema* – Reddening of the mucosa with no discrete lesions present.

Category (1) *Intramucosal petechiae* – A small haemorrhagic lesion lying beneath the mucosal surface.

Category (2) *Haemorrhagic erosion* – A circumscribed mucosal break without depth and with an adherent haemorrhagic clot.

Category (3) *Non-haemorrhagic erosion* – A circumscribed mucosal break without depth and with no adherent clot present.

Category (4) *Superficial ulcer* – A circumscribed mucosal break greater than 3 mm in diameter with definite depth of less than 2 mm.

Category (5) *Deep ulcer* – A circumscribed mucosal break of greater than 3 mm in diameter with depth of greater than 2 mm.

At the end of the recruitment period seven experienced gastroenterologists (three consultant gastroenterologists, two senior registrars, and two post registrar research fellows), each with at least four years experience of upper gastrointestinal endoscopy (average 7·3 years) met to evaluate and categorise independently the lesions, using a structured form based upon the above definitions. All had received training in a wide range of accredited gastroenterology units in British teaching hospitals. At the time of the study four of the observers were fully accredited in gastroenterology and three were at an advanced stage in their accreditation programme. The order in which images were

shown was randomised both for the nature of the lesion and for the patients' drug consumption.

STATISTICAL ANALYSIS

Each lesion was categorised using two methods – firstly, according to the original endoscopist's assessment and, secondly, by the category assigned by the greatest number of observers (consensus category). Statistical analysis was then performed on these two parameters. A mean proportion of agreement (with 95% confidence interval) between observers for each category of lesion was calculated.[11] The Kappa statistic, which measures agreement occurring in excess of that expected by chance, was also calculated.[12]

Results

There was complete agreement that the nine images recorded from non-NSAID patients were ulcers. The consensus view was that one was superficial and eight were deep ulcers. Overall agreement for these subcategories was 86%. By contrast, for NSAID lesions, overall inter-observer agreement was only 52% with a Kappa statistic of 0·37 for the consensus category and 0·38 for the original endoscopists category. Forty four of the NSAID lesions were classified as ulcers but only nine were subclassified as deep ulcers. There were 30 images of non-haemorrhagic erosions and nine of intramucosal haemorrhage. As shown in the Table agreement for individual lesions ranged between only 51% and 60%. Forty two per cent of lesions originally diagnosed as deep ulcers were classified as superficial ulcers; 30% of superficial ulcers were classified as erosions and 17% of erosions were classified as superficial ulcers.

Discussion

The degree of interobserver agreement between endoscopists has important implications for the interpretation of endoscopic studies involving more than one endoscopist or centre, particularly in the assessment of NSAID associated gastric damage. There was complete agreement that the nine lesions seen in patients who were not taking NSAIDs were ulcers. All but one were considered deeper than 2 mm. Similar agreement did not exist for NSAID lesions: the 95% confidence intervals for the proportion of agreement for all types of lesion included values of less than 50%, a level indicative of poor agreement.[11] Ulcers were much more likely to be subclassified as superficial ulcers and cross-classification between ulcers and erosions was widespread.

Few previous studies have addressed the issue of endoscopic interobserver variation.[13–17] Although agreement between two observers seems high for both fixed point scales and visual analogue scores,[13 14] when studies compare agreement between several observers results are less uniform, particularly for qualitative assessment of lesions.[16] Similarly estimation of ulcer size by endoscopists may vary by a factor of eight.[17] No studies, however, have asked the more fundamental question of whether the diagnosis of ulcer itself is a reliable one.

This study was prompted by the observation that the number of ulcers detected endoscopically in patients taking NSAIDs[1–8] was disproportionately high in relation to the three to fourfold magnification of the risk of clinically important end points which epidemiological studies identify as being associated with the use of NSAIDs.[9 10] This study suggests that lesions diagnosed endoscopically as ulcers in NSAID patients are superficial and similar to erosions and may, therefore, have a relatively low probability of progression to clinically important end points such as haematemesis and melaena. While deeper and potentially life threatening ulcers must start as superficial ones, the problem is to identify at endoscopy those which will do so and to define the risk factors and mechanisms by which these are selected. Until this is achieved, uncertainties about the predictive value of ulcers discovered at endoscopy in patients taking NSAIDs raise some doubts about the clinical and prognostic importance of prophylaxis studies which use them as end points. Similarly, confusion in the categorising of NSAID associated gastric erosions may compromise their importance in predicting future ulceration.[18]

Our study was made possible by the use of video endoscopy, since this allows recording of high quality images and subsequent viewing by a panel under controlled blinded conditions. Our panel had been trained in a wide range of units within British gastroenterology. The methodology used is suitable for extension to a larger number of gastroenterologists (for example, by presentation at a national meeting), when evaluation of a wider range of lesions, and subgroup analysis by experience could be investigated. When more valid definitions of gastroduodenal lesions, particularly in patients taking NSAIDs, are agreed video endoscopy could also be used as a vehicle of audit and training to ensure the

Interobserver variation between endoscopists for gastroduodenal mucosal lesions

	Proportion of agreement (95% CI)	
	Original endoscopist category	*Consensus endoscopists category*
Non-NSAID lesions (%):		
Ulcers	100 (100, 100) (n=9)	100 (100, 100) (n=9)
Of which deep ulcers	86 (69, 100) (n=8)	86 (69, 100) (n=8)
NSAID lesions (%):		
Intra-mucosal haemorrhage	60 (45, 76) (n=9)	60 (45, 76) (n=9)
Non-haemorrhagic erosions	60 (49, 99) (n=27)	59 (49, 69) (n=30)
Ulcers	51 (45, 58) (n=47)	51 (44, 57) (n=44)
Of which superficial ulcers	51 (43, 59) (n=33)	51 (44, 57) (n=35)
Of which deep ulcers	51 (40, 63) (n=14)	55 (39, 72) (n=9)
Kappa statistic for NSAID associated lesions	0·38	0·37

Consensus endoscopists:		*Original endoscopist:*	
Deep ulcers:	25·3% Observations classified as superficial ulcers	Deep ulcers:	41·8% Classified as superficial ulcers
Superficial ulcers:	13·1% Classified as deep ulcers 18·8% Classified as non-haemorrhagic erosions	Superficial ulcers:	6·9% Classified as deep ulcers 29·9% Classified as non-haemorrhagic erosions
Non-haemorrhagic erosions:	15·2% Classified as superficial ulcers 1·9% Classified as deep ulcers	Non-haemorrhagic erosions:	16·9% Classified as superficial ulcers 2·6% Diagnosed as deep ulcers

greater accuracy and uniformity of diagnosis that will be required before the importance of a 'NSAID-associated ulcer' discovered at endoscopy can be assessed.

Part of this work was previously presented at the British Gastroenterological Society Meeting, London, September 1991.

1 Silvoso GR, Ivey KJ, Butt JH, *et al.* Incidence of gastric lesions in patients with rheumatic disease on chronic aspirin therapy. *Ann Intern Med* 1979; **91:** 517–20.
2 Morris AD, Holt SD, Silvoso GR, *et al.* Effect of anti-inflammatory drug administration in patients with rhematoid arthritis. *Scand J Gastroenterol* 1981; **16** (suppl 67): 131–5.
3 Lockhard O, Ivey KJ, Butt JH, Silvoso GR, Sisk K, Holt S. The prevalence of duodenal lesions in patients with rheumatic diseases on chronic aspirin therapy. *Gastrointest Endosc* 1980; **26:** 5–7.
4 Collins AJ, Du Toit JA. Upper gastrointestinal findings and faecal occult blood in patients with rheumatic diseases taking nonsteroidal anti-inflammatory drugs. *Br J Rheumatol* 1987; **26:** 295-8.
5 Larkai EN, Smith LJ, Lidsky MD, Graham DY. Gastroduodenal mucosa and dyspeptic symptoms in arthritic patients during chronic nonsteroidal anti-inflammatory drug use. *Am J Gastroenterol* 1987; **82:** 1153–8.
6 Yeomans ND, Elliot SL, Edwards J, Buchanan R, Sturrock SD, Smallwood R. Gastroduodenal damage during therapy with non stroidal anti-inflammatory drugs. *Gastroenterology* 1988; **94:** A510.
7 Biewer W, Buschmeier B, Bolten W. Risk factors of gastrointestinal injury: Prospective multicentre study on 497 rheumatic patients. *Gut* 1990; **31:** A1176.
8 Geis S, Stead H, Wallemark C-B, Nicholson PA. Prevalence of mucosal lesions in the stomach and duodenum due to chronic use of NSAID in patients with rheumatoid arthritis or osteoarthritis, and interim report on prevention by misoprostol of diclofenac associated lesions. *J Rheumatol* 1991; **18** (suppl 28): 11–4.
9 Hawkey CJ. Non-steroidal anti-inflammatory drugs and peptic ulcers: facts and figures multiply but do they add up? *BMJ* 1990; **300:** 278–84.
10 Langman MJS, Brooks P, Hawkey CJ, Silverstein F, Yeomans N. Non steroidal anti-inflammatory drug associated ulcer: epidemiology causation and treatment. *J Gastroenterol Hepatol* 1991; **6:** 442–9.
11 Grant JM. The fetal heart rate is normal, isn't it? Observer agreement of categorical assessments. *Lancet* 1991; **337:** 215–8.
12 Siegel S, Castellan NJ. Measures of association and their tests of significance. In: *Non parametric statistics for behavioural sciences.* 2nd Ed. New York: McGraw-Hill, 1988.
13 Aabakken L, Larsen S, Osnes M. Visual analogue scales for endoscopic evaluation of nonsteroidal anti-inflammatory drug induced mucosal damage in the stomach and duodenum. *Scand J Gastroenterol* 1990; **25:** 443–8.
14 Lanza FL, Royer GL, Nelson RS, Chen TT, Seckman SE, Rack MF. The effects of ibuprofen, indomethacin, aspirin, naproxen and placebo on the gastric mucosa of healthy volunteers. A gastroscopic and photographic study. *Dig Dis Sci* 1979; **24:** 823–8.
15 Hawthorne AB, Hurst SM, Mahida YR, Cole AT, Hawkey CJ. Aspirin-induced gastric mucosal damage: prevention by enteric coating of aspirin and relation to prostaglandin synthesis. *Br J Clin Pharmacol* 1990; **30:** 187–94.
16 Cales P, Zabotto B, Meskens C, Caucanas J-P, Vinel J-P, Desmorat H, *et al.* Gastroesophageal endoscopic features in cirrhosis, observer variability, interassociations and relationship to hepatic dysfunction. *Gastroenterology* 1990; **98:** 156–62.
17 Sonnenberg A, Giger M, Kern L, Noll C, Stuby K, Weber KB, Blum AL. How reliable is determination of ulcer size by endoscopy. *BMJ* 1979; **2:** 1322–4.
18 Agrawal N, Meckstroth S, Dajani E. Clinical significance of gastric erosions as a predictor of NSAID-induced ulcers. *Gastroenterology* 1992; **104:** A1.