Inter-endoscopist agreement in diagnosis of Barrett's oesophagus

Glen A Doherty,¹ Danny G Cheriyan,¹ Jan E Leyden,¹ John F O'Dowd,² Frank E Murray,¹ Stephen E Patchett¹

► Supplementary figure 1 is published online only. To view this file please visit the journal online (<http://fg.bmj.com>).

¹Department of Gastroenterology, Beaumont Hospital/Royal College of Surgeons in Ireland, Dublin, Ireland

²Department of Pathology, Bon Secours Hospital/Royal College of Surgeons in Ireland, Dublin, Ireland

Correspondence to

Dr S Patchett, Department of Gastroenterology, Beaumont Hospital/Royal College of Surgeons in Ireland, Beaumont Road, Dublin 9, Ireland; patchett@iol.ie

Accepted 13 December 2010
Published Online First
20 March 2011

Objective To assess how interpretation of abnormalities at the oesophago-gastric junction (OGJ) when making a diagnosis of Barrett's oesophagus (BO) varies between endoscopists and to examine the impact of the endoscopy experience on these decisions.

Design/setting Members of the Irish Society of Gastroenterology who regularly perform gastroscopy were invited to participate in a web based image assessment study.

Interventions Questions were posed to ascertain level of endoscopy experience, and participants were asked to indicate the presence or absence of BO in 12 endoscopic images of the OGJ.

Outcome measures Primary outcome was overall level of agreement in responses and relationship to endoscopy experience.

Results The responses of 65 clinicians regularly performing gastroscopy were analysed. In 3/12 images, showing typical long segment BO, there was a strong consensus on the endoscopic diagnosis (>95% agreement). However, agreement was fair to poor (κ for multiple raters, 0.31) on the presence or absence of short BO segments at endoscopy. Minimal differences were observed between experienced endoscopists (individuals with >10 years' endoscopy experience) and less experienced counterparts in the threshold for BO diagnosis. Inter-endoscopist agreement overall was not significantly better within the more experienced group.

Conclusion The study demonstrates low interobserver agreement in endoscopic diagnosis of (short segment) BO, even among experienced endoscopists. Given the costs associated with endoscopic surveillance of BO, prompt efforts to promote consensus diagnosis and improve agreement are required as an important quality improvement measure in this area.

Introduction

Evidence of a consistent rise in the incidence of oesophageal adenocarcinoma over recent decades^{1–3} has focused interest on Barrett's (columnar lined) oesophagus (BO), the precursor lesion

whose incidence also appears to be on the increase.^{4–5} The association between BO and cancer has led to the development of recommendations for interval endoscopic screening with the goal of early detection of dysplasia.^{6–8} However, screening comes at a cost, with both physical and psychological morbidity and financial implications both for the individual and health system.⁹

The diagnosis of BO relies on visualisation at endoscopy of columnar mucosa above the oesophago-gastric junction (OGJ), with confirmation or corroboration on biopsy samples. Making a confident and correct initial endoscopic diagnosis is clearly important in minimising long term costs and morbidity, although relatively little attention has been given to the quality and reproducibility of endoscopic findings of BO. Those studies that have looked at the area have raised concerns about the operating characteristics of endoscopy in diagnosing BO¹⁰ and demonstrate limitations in the reproducibility of reported endoscopic findings in a significant minority of patients.¹¹ There have also been suggestions that the level of endoscopy experience has an impact on the quality of endoscopic diagnosis of BO.¹²

In cases where the endoscopist is uncertain about the presence or absence of BO, biopsy findings typically do not add clarity as definitive histological evidence of BO is present in only a small percentage of patients.^{13–14} While the presence of intestinal metaplasia (IM) may corroborate the diagnosis, reliance on the absence of IM on index biopsy neither excludes the diagnosis nor offers much reassurance to the patient in terms of cancer risk.¹⁵ The endoscopist can typically expect to receive confirmation of the presence of non-intestinal columnar epithelium (the significance of which depends entirely on the biopsy site). Thus given the importance

of the interpretation of initial endoscopic appearances, we set out to examine the levels of agreement between endoscopists about the presence of endoscopic appearances of BO and sought to evaluate the potential importance of endoscopy experience in determining the threshold for making this endoscopic diagnosis.

Methods

Internet based image assessment study

Invitations to participate in an internet based image assessment study were distributed via email to members of the Irish Society of Gastroenterology (ISG). The ISG membership comprises gastroenterologists in practice and in training, surgeons, other clinicians and researchers with an interest in gastroenterology. The ISG does not issue or endorse specific clinical guidelines on the diagnosis of BO but local practice in Ireland is generally in line with the British Society of Gastroenterology recommendations.¹⁶ The email contained a link to a web page which featured a range of questions to ascertain the nature of practice of participating clinicians and their level of endoscopy experience and activity. Twelve selected anonymised static endoscopic images of the OGJ were then available to view on the webpage (taken from patients with a history of reflux symptoms or BO who underwent endoscopy at Beaumont Hospital, Dublin, Ireland during a 6 week period in November and December 2007). Participants were asked to select responses on the web page to indicate an endoscopic diagnosis of 'normal' or 'BO' for each image. When satisfied with their answers, they were invited to hit a 'submit' button which returned the results to an electronic database for analysis. No identifiable electronic link to the identity of participants was stored. The study was performed with the approval of Beaumont Hospital Ethics Committee.

Statistical analysis

Comparison of differences in proportions was performed using χ^2 tests, performed using Instat/GraphPad Prism Software. Kappa statistics for agreement between multiple raters (for nominal data) were performed according to the method described by Siegel and Castellan.¹⁷ A p value <0.05 was considered significant.

Results

Seventy-three individuals participated in the study, 65 of whom reported regularly performing gastroscopy as part of their clinical practice; only these responses were analysed. Most participants were physicians ($n=57$), with a small number of responses from surgeons. The combined participation rate for gastroenterology consultant and specialist registrar (SpR) members of the ISG was 51% (49/97). Approximately half of the respondents were practicing as consultants (32/65, 49%), the remainder working as either registrars or SpRs. Almost half of participants (29/65,

45%) had more than 10 years' endoscopy experience and two-thirds (43/65) reported performing over 250 gastroscopies per annum. The participants viewed 12 digital images of the OGJ and were asked to make a judgement on whether the appearances were normal or consistent with BO. The 12 images can be viewed in the supplementary file (figure S1, available online only; images are numbered P1–P12, consistent with the order in which they were presented).

There was a high level of agreement between participants ($>95\%$ responses) on the presence of endoscopic appearances consistent with BO in three of the 12 images presented (P1, P7, P10). These arguably show 'classic' appearances of a long Barrett's segment (>3 cm), with either the presence of a wide area of columnar epithelium visible above the confluence of the gastric folds or islands of squamous mucosa within a columnar segment. Levels of agreement were less for the remaining nine images; the responses are summarised in figure 1. Overall, more than 15% of the participants disagreed with the preferred diagnosis of the group in two-thirds (8/12) of the images. In one case (P9), the responses were split almost equally between those who supported a diagnosis of BO and those calling the appearances 'normal'. In others, a substantial minority disagreed with the preferred diagnosis. The overall level of concordance between participants across the 12 images was fair to poor, with κ statistic for multiple raters of 0.31 ($z=13.92$, $p<0.001$).

Table 1 summarises the findings relating to the level of endoscopy experience; according to consultant versus trainee status, level of endoscopy experience and endoscopy volume for the three images with the highest level of disagreement in responses (where differences related to these factors ought to be most visible). No significant differences were observed in the proportion of consultants (vs trainees), of experienced endoscopists (vs those with <10 years of experience) or of higher volume practitioners (vs those performing less than 250/year) favouring a diagnosis of BO in their responses to images P4 or P9. χ^2 statistics reflecting significant differences between groups in responses to image P8 were related to the high 'non-response' rate in the consultant/experienced endoscopist groups. When non-responders were excluded, these differences were not significant.

Finally, with the aim of further exploring this question, all responses were analysed to ascertain whether the level of agreement was different between individuals *within* the same group (consultant or trainee endoscopist). The κ statistic for the level of agreement in responses (across all 12 images) within the 'consultants' group was $\kappa=0.32$ ($z=9.81$, $p<0.001$) and showed no difference from the level of agreement for responses within 'trainees' taken together as a group ($\kappa=0.29$; $z=8.85$, $p<0.001$). These findings again suggested that level of endoscopy experience was not a significant factor in determining response.

Figure 1 Responses of participants (when asked to indicate the presence or absence of Barrett's oesophagus (BO)) to all 12 images. Each bar represents the responses to a single image; the numbers of the images are those shown in supplementary figure S1 (available online only) and are numbered in the order in which the images were presented. The bars show the proportion of responses which favoured a diagnosis of BO (black) and normal (white). Non-responses for each image are shown in grey.

Discussion

BO (columnar lined) is found in 5–10% of patients with chronic gastro-oesophageal reflux symptoms attending for endoscopy.^{18–20} The greatest diagnostic accuracy arises from a combination of visualisation of epithelial columnarisation above the OGJ and histological confirmation or corroboration of this finding.¹⁶ While demonstration of the presence of IM on biopsy has previously been considered a necessary part of a definitive diagnosis in certain countries, such as the USA,⁷ opinion is rapidly shifting closer to the framework accepted in the UK and Europe.²¹ Sampling variations means that failure to demonstrate IM on biopsy does not reliably exclude BO.¹⁶ Indeed, it is increasingly recognised that in BO, columnar epithelium without IM is biologically intestinalised²² and shows similar propensity to neoplastic transformation.^{15–23} Conversely, IM can be seen in biopsies from the cardia and does not appear to confer the risk of dysplasia and cancer associated with even short segment BO.^{24–25}

All of this shifts the emphasis back to a correct interpretation of the initial endoscopic findings. Biopsy findings generally only serve as adjunctive to a primary endoscopic diagnosis. Difficulties may arise in the correct endoscopic identification of BO, particularly in the classification of minor abnormalities at the OGJ which may represent a short BO segment. Despite calls to standardise the endoscopic criteria used to describe the OGJ²⁶ and the way we define and diagnose short segment BO,²⁵ there has been little effort made to critically evaluate 'real life' differences in how endoscopists report more subtle endoscopic abnormalities. Undercalling these findings may mean 'at risk' patients are missed and overcalls lead to unnecessary worry, expense or an erroneous initiation of surveillance (which, once started, can be difficult to stop).

This study examined this issue by evaluating levels of agreement between participants in making an 'endoscopic' diagnosis of BO using an internet based image assessment tool and looking at the impact, if any, of endoscopy experience on the threshold for making a diagnosis. While levels of agreement between participants were excellent when evaluating images of classic (long segment, >3 cm) BO, levels of inter-endoscopist agreement overall were only poor to fair. While use of a validated endoscopic grading system (Prague C&M criteria) for BO can produce impressive levels of agreement regarding the extent of disease,²⁷ even expert endoscopists disagree in their interpretation of minor abnormalities at the OGJ. The κ values for agreement between expert endoscopists on BO segments <1 cm in length in the Prague study was reported as 0.21.²⁷ This is broadly in line with our observations of agreement between 'non-expert' endoscopists, where there was poor consensus in interpretation of similar 'ultra-short' lesions. It is clear that what one endoscopist calls an 'irregular Z line' represents a sufficient abnormality for others to raise the Barrett's flag. Despite some previous studies suggesting that level of endoscopy experience has an impact on diagnostic accuracy for BO,¹² our results do not demonstrate any greater agreement among consultant level/experienced endoscopists (>10 years of practice) than their less experienced counterparts. Instead, our results reflect the findings of others, demonstrating low levels of inter-endoscopist agreement in identification of the OGJ, irrespective of experience and specialist training.²⁸

There are weaknesses in this study, the principal one being that viewing a single still image may not equate to a dynamic 'real time' endoscopic evaluation of the OGJ. However, we observed a low rate

Table 1 The responses of participants to images P4, P8 and P9

	Barrett's	Normal	Blank	χ^2 (df=2)
Response (P4)
Trainee	21 (0.66)	10 (0.31)	1 (0.03)	1.40
Consultant	20 (0.61)	13 (0.39)	0 (0)	p=0.49
≤250 procedures/year	12 (0.55)	9 (0.40)	1 (0.05)	2.63
>250 procedures/year	29 (0.67)	14 (0.33)	0 (0)	p=0.26
≤10 years endoscopy experience	23 (0.64)	12 (0.33)	1 (0.03)	0.90
>10 years endoscopy experience	18 (0.62)	11 (0.38)	0 (0)	p=0.63
Response (P8)
Trainee	21 (0.66)	11 (0.34)	0 (0)	5.87
Consultant	21 (0.64)	7 (0.21)	5 (0.15)	p=0.05*
≤250 procedures/year	15 (0.68)	6 (0.27)	1 (0.05)	0.49
>250 procedures/year	27 (0.63)	12 (0.28)	4 (0.09)	p=0.78
≤10 years endoscopy experience	24 (0.67)	12 (0.33)	0 (0)	7.17
>10 years endoscopy experience	18 (0.62)	6 (0.21)	5 (0.17)	p=0.02*
Response (P9)
Trainee	16 (0.50)	16 (0.50)	0 (0)	1.05
Consultant	15 (0.45)	17 (0.52)	1 (0.03)	p=0.59
≤250 procedures/year	12 (0.55)	10 (0.45)	0 (0)	1.02
>250 procedures/year	19 (0.44)	23 (0.54)	1 (0.02)	p=0.60
≤10 years endoscopy experience	18 (0.50)	18 (0.50)	0 (0)	1.34
>10 years endoscopy experience	13 (0.45)	15 (0.52)	1 (0.03)	p=0.51

Small corresponding digital images are included within the table. The table shows the number of participants (with corresponding proportions in parentheses) favouring either a diagnosis of Barrett's oesophagus, 'normal' or offering no response (blank) within the groups, as shown (eg, consultant vs trainee, etc). Also shown is the χ^2 statistics (with 2 degrees of freedom) for the comparison of responses between each paired group, with the corresponding p value shown below in each case.

of non-responses (where participants failed to offer a 'diagnosis' for a particular image), suggesting that both the image quality was adequate and there was adequate 'visual' information within the static image for a judgement to be made. We have been exploring using short endoscopic video sequences but difficulties with download speeds and image quality led us to proceed with static images in the first instance. We envisage validating our findings in the future using a web

based system to allow participants to view dynamic video footage. The participation rate in this study among gastroenterology consultants and SpRs (the single largest group within the clinical membership of the ISG) exceeded 50%. While a higher participation rate would improve the generalisability of the results, we believe the rate achieved is broadly comparable with other similar studies and does not impact on the substantive findings.

The challenge which emerges from our results is how to improve inter-endoscopist agreement in the interpretation of subtle abnormalities at the OGJ. While endoscopy experience does not appear to be important, there may be a role for education. While complex endoscopic classification systems are unpopular for use in day to day clinical practice (outside of clinical trials), an improved awareness of existing systems to describe the Z line (such as the ZAP classification) may be useful in selecting those patients more likely to have IM at biopsy.²⁸ Education about these criteria does seem to improve the level of agreement in how endoscopists report appearances at the OGJ.^{28–29} In the near future, so called ‘multi-modal’ endoscopic imaging of the oesophagus offers the promise of better visualisation and better targeting of biopsies,³⁰ but with the risk that endoscopists may also be left with more things to disagree about. Our own approach has been to encourage extensive photography or video imaging by endoscopists of abnormalities at the OGJ for review with colleagues, ideally in the context of a multidisciplinary team meeting with a consensus diagnosis and management plan arrived at following discussion between endoscopists (as a group) and histopathologists. It has been suggested that columnar segments less <1 cm in length should neither be reported nor biopsied,²¹ given the lack of diagnostic certainty, low risk of dysplasia and doubtful benefit of surveillance. This is a view to which we are sympathetic and a strong endorsement of such an approach by future national and international practice guidelines might improve the confidence of endoscopists to adopt this approach.

In combination with other measures to improve the quality of diagnostic practice in BO,^{31–32} harm can be minimised, patient care enhanced and scarce endoscopy resources used to good effect (with avoidance of inappropriate surveillance procedures). However, it is only with the recognition that we may not all ‘see the same things’ that we are likely to improve endoscopy reporting quality in this area.

Acknowledgements The authors acknowledge the assistance of Mr Michael Dineen, CEO of the Irish Society of Gastroenterology, and of Mr Richard O’Connor, of Curratech Internet Consultancy.

Competing interests None.

Ethics approval The study was conducted with the approval of the Beaumont Hospital Ethics (Medical Research) Committee, Dublin, Ireland.

Provenance and peer review Not commissioned; externally peer reviewed.

References

1. Blot WJ, Devesa SS, Kneller RW, *et al.* Rising incidence of adenocarcinoma of the esophagus and gastric cardia. *JAMA* 1991;265:1287–9.
2. Devesa SS, Blot WJ, Fraumeni JF Jr. Changing patterns in the incidence of esophageal and gastric carcinoma in the United States. *Cancer* 1998;83:2049–53.
3. Powell J, McConkey CC, Gillison EW, *et al.* Continuing rising trend in oesophageal adenocarcinoma. *Int J Cancer* 2002;102:422–7.
4. Caygill CP, Reed PI, Johnston BJ, *et al.* A single centre’s 20 years’ experience of columnar-lined (Barrett’s) oesophagus diagnosis. *Eur J Gastroenterol Hepatol* 1999;11:1355–8.
5. Prach AT, MacDonald TA, Hopwood DA, *et al.* Increasing incidence of Barrett’s oesophagus: education, enthusiasm, or epidemiology? *Lancet* 1997;350:933.
6. Sampliner RE. Practice guidelines on the diagnosis, surveillance, and therapy of Barrett’s esophagus. The Practice Parameters Committee of the American College of Gastroenterology. *Am J Gastroenterol* 1998;93:1028–32.
7. Wang KK, Sampliner RE. Updated guidelines 2008 for the diagnosis, surveillance and therapy of Barrett’s esophagus. *Am J Gastroenterol* 2008;103:788–97.
8. Wang KK, Wongkeesong M, Buttar NS. American Gastroenterological Association medical position statement: role of the gastroenterologist in the management of esophageal carcinoma. *Gastroenterology* 2005;128:1468–70.
9. Wright TA, Gray MR, Morris AI, *et al.* Cost effectiveness of detecting Barrett’s cancer. *Gut* 1996;39:574–9.
10. Eloubeidi MA, Provenzale D. Does this patient have Barrett’s esophagus? The utility of predicting Barrett’s esophagus at the index endoscopy. *Am J Gastroenterol* 1999;94:937–43.
11. Kim SL, Waring JP, Spechler SJ, *et al.* Diagnostic inconsistencies in Barrett’s esophagus. Department of Veterans Affairs Gastroesophageal Reflux Study Group. *Gastroenterology* 1994;107:945–9.
12. Padua S, Ramirez FC. Accuracy in the diagnosis of short-segment Barrett’s esophagus: the role of endoscopic experience. *Gastrointest Endosc* 2001;54:605–8.
13. Takubo K, Nixon JM, Jass JR. Ducts of esophageal glands proper and paneth cells in Barrett’s esophagus: frequency in biopsy specimens. *Pathology* 1995;27:315–17.
14. Srivastava A, Odze RD, Lauwers GY, *et al.* Morphologic features are useful in distinguishing Barrett esophagus from carditis with intestinal metaplasia. *Am J Surg Pathol* 2007;31:1733–41.
15. Kely CJ, Gough MD, Van Wyk Q, *et al.* Barrett’s oesophagus: intestinal metaplasia is not essential for cancer risk. *Scand J Gastroenterol* 2007;42:1271–4.
16. Playford RJ. New British Society of Gastroenterology (BSG) guidelines for the diagnosis and management of Barrett’s oesophagus. *Gut* 2006;55:442.
17. Siegel S, Castellan NJ Jr. *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill Book Company, 1988.
18. Winters C Jr, Spurling TJ, Chobanian SJ, *et al.* Barrett’s esophagus. A prevalent, occult complication of gastroesophageal reflux disease. *Gastroenterology* 1987;92:118–24.
19. Mann NS, Tsai MF, Nair PK. Barrett’s esophagus in patients with symptomatic reflux esophagitis. *Am J Gastroenterol* 1989;84:1494–6.
20. Corder AP, Jones RH, Sadler GH, *et al.* Heartburn, oesophagitis and Barrett’s oesophagus in self-medicating patients in general practice. *Br J Clin Pract* 1996;50:245–8.
21. Riddell RH, Odze RD. Definition of Barrett’s esophagus: time for a rethink—is intestinal metaplasia dead? *Am J Gastroenterol* 2009;104:2588–94.

22. **Hahn HP**, Blount PL, Ayub K, *et al.* Intestinal differentiation in metaplastic, nongoblet columnar epithelium in the esophagus. *Am J Surg Pathol* 2009;33:1006–15.
23. **Gatenby PA**, Ramus JR, Caygill CP, *et al.* Relevance of the detection of intestinal metaplasia in non-dysplastic columnar-lined oesophagus. *Scand J Gastroenterol* 2008;43:524–30.
24. **Sharma P**, Weston AP, Morales T, *et al.* Relative risk of dysplasia for patients with intestinal metaplasia in the distal oesophagus and in the gastric cardia. *Gut* 2000;46:9–13.
25. **Sharma P**, Morales TG, Bhattacharyya A, *et al.* Dysplasia in short-segment Barrett's esophagus: a prospective 3-year follow-up. *Am J Gastroenterol* 1997;92:2012–16.
26. **Armstrong D.** Review article: towards consistency in the endoscopic diagnosis of Barrett's oesophagus and columnar metaplasia. *Aliment Pharmacol Ther* 2004;20(Suppl 5):40–7.
27. **Sharma P**, Dent J, Armstrong D, *et al.* The development and validation of an endoscopic grading system for Barrett's esophagus: the Prague C & M criteria. *Gastroenterology* 2006;131:1392–9.
28. **Amano Y**, Ishimura N, Furuta K, *et al.* Which landmark results in a more consistent diagnosis of Barrett's esophagus, the gastric folds or the palisade vessels? *Gastrointest Endosc* 2006;64:206–11.
29. **Wallner B**, Sylvan A, Janunger KG. Endoscopic assessment of the 'Z-line' (squamocolumnar junction) appearance: reproducibility of the ZAP classification among endoscopists. *Gastrointest Endosc* 2002;55:65–9.
30. **Sharma P**, Bansal A. Toward better imaging of Barrett's esophagus—see more, biopsy less! *Gastrointest Endosc* 2006;64:188–92.
31. **Ofman JJ**, Shaheen NJ, Desai AA, *et al.* The quality of care in Barrett's esophagus: endoscopist and pathologist practices. *Am J Gastroenterol* 2001;96:876–81.
32. **Curvers WL**, Peters FP, Elzer B, *et al.* Quality of Barrett's surveillance in The Netherlands: a standardized review of endoscopy and pathology reports. *Eur J Gastroenterol Hepatol* 2008;20:601–7.