

Variations in the reporting of endoscopies by different endoscopists

HL Spencer, AJ Lobo and SA Riley

ABSTRACT – All gastroscopies and colonoscopies performed in two UK teaching hospitals over a period of one year were audited to investigate whether endoscopic reporting of gastroscopies and colonoscopies by different endoscopists is consistent. Endoscopic diagnoses were retrieved from the hospitals' endoscopy databases. The results of 1,814 colonoscopies and 2,127 gastroscopies were analysed using χ^2 (Chi squared). The frequency of reporting common diagnoses was variable and the differences between specialist endoscopists were highly significant, including for important conditions such as peptic ulceration (range 2–10%, $p=0.001$) and colonic polyps (16–45%, $p<0.0001$). There is a large variation in the frequency of the diagnoses reported by different endoscopists. This is unlikely to be explained by casemix or chance. This may have major implications for the health of patients. More emphasis must be placed during training on the correct interpretation of endoscopies.

KEY WORDS: colonoscopy, diagnosis, endoscopy, gastroscopy, interpretation, reporting, training

Introduction

Endoscopic examination is a fundamental investigation in the armoury of the gastroenterologist. In the UK alone almost one million gastrointestinal endoscopies are carried out every year.¹ Endoscopy allows direct visualisation of internal organs. Although biopsy to confirm diagnosis is possible, an immediate final diagnosis is made in most cases. This makes it essential that endoscopic interpretation of macroscopic appearances is accurate and consistent between different endoscopists. However, training in endoscopic interpretation is variable and most studies suggest large variations in diagnostic consistency even between expert endoscopists.

There is evidence that both early gastric cancer and colonic cancer may not be detected due to inadequate examination.^{2,3} Careful examination is therefore vital if subtle but serious lesions are not to be missed. In addition, over-interpretation of findings may also be a problem. Endoscopic findings such as

gastritis may correlate poorly with histology,⁴ resulting in an inappropriate clinical label.

There is increasing interest in the quality of procedures,^{5,6} and the difficulties of determining competence.^{7,8} Guidelines drawn up by the American Society of Gastrointestinal Endoscopy (ASGE) and the British Society of Gastroenterology (BSG) emphasise technical competence and total numbers of procedures completed. The importance of interpretive skills is only briefly discussed and evaluation of these skills is not standardised.^{9,10}

There is little published literature regarding diagnostic consistency among endoscopists.^{11,12,13} Some evidence suggests that there may be considerable variability in interpreting endoscopic appearances, with potentially important clinical consequences.^{14,15} Many of the published studies concentrate on 'expert' endoscopists looking at relatively small numbers of patients. The specific aim of this study was, therefore, to study the variability of endoscopic reporting in general clinical practice.

Methods

Two hospital databases were analysed for a single calendar year. The data was collected prospectively. The endoscopists completed a single, final report immediately following every procedure. The two hospitals, Royal Hallamshire Hospital (RHH) and Northern General Hospital (NGH), are teaching hospitals in Sheffield.

Endoscopists were unaware that the study would be carried out, but were aware that the databases could be audited.

At RHH, all gastroscopies performed on a joint medical waiting list between 1 January 2000 and 31 December 2000 were recorded on an Endoscribe database and subsequently analysed. At NGH, data on colonoscopies was recorded on a customised Infoflex system between 1 October 2001 and 30 September 2002. The colonoscopic procedures were listed from a joint medical and surgical list.

The first author worked at both institutions during different periods of time; hence the data for gastroscopy (RHH) and colonoscopy (NGH) were collected independently and over two separate periods. With the exception of the first author, the

HL Spencer MRCP, Consultant Physician and Gastroenterologist, Chesterfield Royal Hospital, Calow, Chesterfield

AJ Lobo FRCP, Consultant Physician and Gastroenterologist, Royal Hallamshire Hospital, Sheffield

SA Riley FRCP, Consultant Physician and Gastroenterologist, Northern General Hospital, Sheffield

Clin Med
2007;7:23–27

endoscopists carrying out gastroscopies and colonoscopies at the two institutions were different.

Emergency and inpatient procedures were not included in either analysis. Certain endoscopists carried out no inpatient procedures; therefore, all inpatient procedures were excluded to ensure a homogenous patient group and prevent selection bias. In addition, endoscopists who had performed fewer than 20 procedures were excluded from the analysis in order to permit meaningful comparisons. One registrar studying coeliac disease had research lists to carry out small bowel biopsies. All of his procedures were excluded from the analysis.

The primary aim was to determine whether there were significant differences in the final diagnoses reported by different specialist endoscopists. A secondary aim was to investigate whether there were differences in endoscopic reporting between specialists and trainees.

Results were analysed using χ^2 to determine whether the frequency of reporting a diagnosis differed significantly between

individual specialist endoscopists. χ^2 was also used to determine whether differences existed in the reporting of various diagnoses between trainees and specialists.

Results

Gastroscopies

Of the 2,127 gastroscopies analysed (985 male, 1,142 female), 576 (27%) were sedated. Two GP specialists, one consultant and five specialist registrars (trainees) conducted the gastroscopies. Trainees carried out 585 (28%) of the gastroscopies. None of the endoscopists had performed fewer than 20 examinations.

The frequency of reporting the common diagnoses from gastroscopies varied greatly and the differences between specialist endoscopists were highly significant. In particular, significant differences were found in the frequency of reporting peptic ulceration ($p=0.001$) and oesophagitis ($p<0.0001$). Differences in the reporting frequency for cancer did not reach statistical significance ($p=0.17$) (Table 1). In addition, individual endoscopists' reporting rates deviated significantly from the mean (Fig 1). The frequency of gastroscopies reported as normal ranged from 26–59% ($p<0.0001$).

Endoscopists carried out between 19% and 41% of gastroscopies under sedation ($p<0.001$).

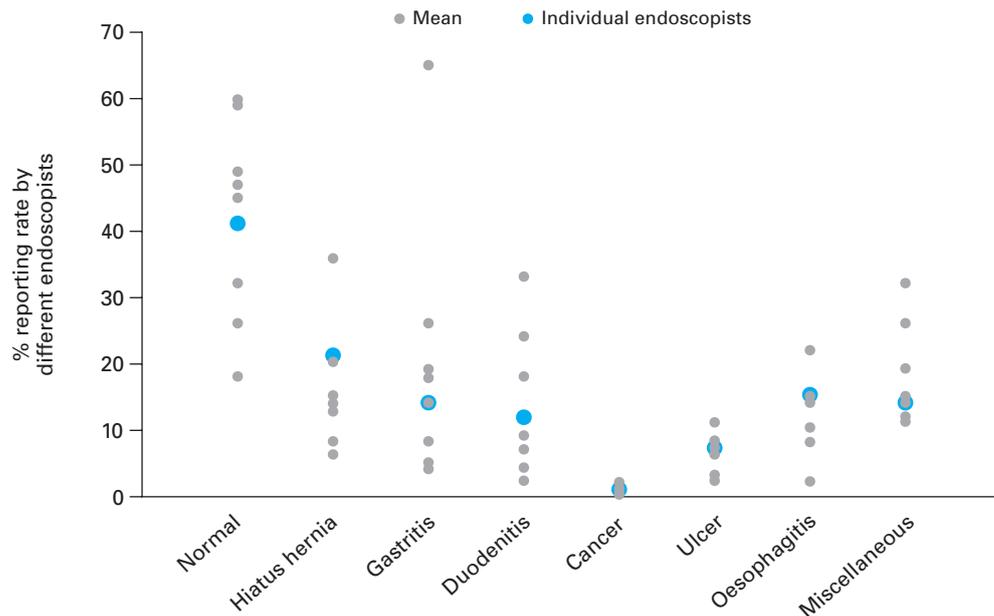
Colonoscopies

During the study period, 1,814 colonoscopies were performed. Of these, 1,776 colonoscopies (796 male, 980 female) were performed by four surgical consultants, four medical consultants, six specialist registrars and two specialist nurse colonoscopists; trainees carried out 267 (15%). Eleven (0.6%) were performed without sedation. The remaining 38 colonoscopies were

Table 1. Gastroscopy reporting by different specialist endoscopists.

Endoscopic diagnosis	Overall total (%) in brackets)	Range between endoscopists (%)	p value
Normal	596 (39)	26–59	$p<0.0001$
Hiatus hernia	381 (25)	14–36	$p<0.0001$
Gastritis	220 (14)	4–26	$p<0.0001$
Duodenitis	263 (14)	2–24	$p<0.0001$
Oesophagitis	261 (17)	8–22	$p<0.0001$
Peptic ulceration	121 (8)	2–10	$p=0.001$
Cancer	12 (1)	0.5–1	$p=0.17$

Fig 1. Variation in gastroscopists (specialists and trainees) reporting from the mean.



performed by five endoscopists who had each performed fewer than 20 procedures and were excluded from analysis.

There was marked variability in the frequency of reporting the common colonoscopic diagnoses and the differences between specialist endoscopists were highly significant. The rates of reporting colonic polyps ($p<0.0001$) and inflammation ($p<0.0001$), but not cancer ($p=0.44$), showed a significant variability between endoscopists (Table 2). In addition, individual endoscopists' reporting rates deviated significantly from the mean (Fig 2). The frequency of colonoscopies reported by different endoscopists to be normal, ranged from 23–55% ($p<0.0001$).

On an intention-to-treat analysis, colonoscopy completion rate was 82% overall (range 64–90%), ($p<0.0001$ for the difference between endoscopists).

General results

The frequency of some diagnoses did vary according to whether the endoscopist was a trainee or a specialist (Table 3). Trainees were more likely to report a normal examination in both gastroscopy ($p<0.0001$) and colonoscopy ($p<0.01$). Trainees had significantly lower reporting rates for peptic ulceration ($p<0.05$), but the reporting of colonic polyps was similar in both groups ($p=0.58$).

Miscellaneous cases included several relatively less common diagnoses including benign oesophageal strictures, varices and leiomyomas. The small numbers of these patients make it impossible to draw statistically relevant conclusions about the individual diagnoses.

Discussion

The present study highlights current inconsistencies in endoscopic diagnosis and reporting. Our study found significant differences for clinically important conditions such as peptic ulcer-

Table 2. Colonoscopy reporting by different specialist endoscopists.

Endoscopic diagnosis	Overall total (%) in brackets	Range between endoscopists (%)	p value
Normal	516 (34)	23–55	$p<0.0001$
Colonic polyps	465 (31)	16–45	$p<0.0001$
Inflammation	230 (15)	4–28	$p<0.0001$
Haemorrhoids	256 (17)	1–35	$p<0.0001$
Diverticulosis	442 (29)	20–38	$p<0.005$
Colonic cancer	61 (4)	1–8	$p=0.44$

ation, oesophagitis and colonic polyps. Both over- and under-reporting may have a detrimental effect on the patient, and the large differences between endoscopists are a cause for concern. Reassuringly, variability in cancer reporting was not statistically significant, but this may reflect the fact that the total number of cancer cases was small (16 gastro-oesophageal and 73 colonic).

There are a small number of studies in the literature examining variability in endoscopic reporting. Some studies have been 'head to head' comparisons of expert endoscopists, involving small numbers of consecutive procedures on the same patient.¹³ Alternatively, some groups have investigated inter-observer agreement using still photographic images.^{11,12,16,17} However, the quality of the stored still images is a significant limitation and resolution is often suboptimal. Endoscopic video recording overcomes this to an extent.^{14,18,19} However, analysis of large numbers of recordings is time consuming. Further, as with stills, a recording cannot emulate the sensitivity of a live endoscopy since it does not allow 'exploration' of the upper or lower gastrointestinal tract in order to clarify appearances.

Fig 2. Variation in colonoscopists (specialists and trainees) reporting from the mean.

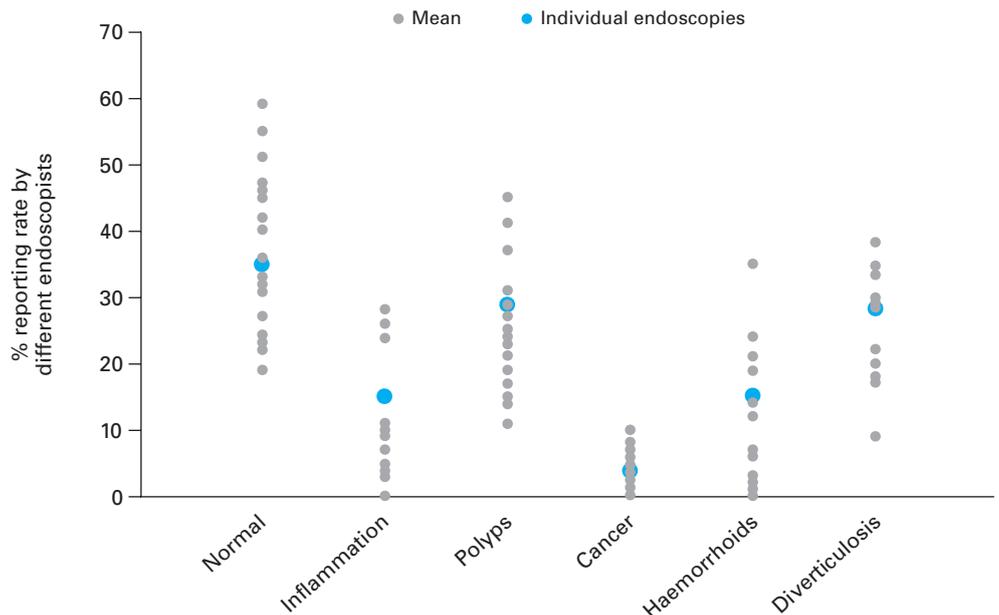


Table 3. Frequency of reporting selected diagnoses comparing trainees and specialists.

Diagnosis	Trainees total (%) in brackets)	Specialists total (%) in brackets)	p value for difference between trainees and specialists
Total gastroscopies	585 (28)	1542 (72)	NA
Normal gastroscopy	284 (49)	596 (39)	p<0.0001
Gastritis	103 (18)	220 (14)	p=0.06
Oesophagitis	49 (8)	261 (17)	p<0.0001
Peptic ulceration	32 (5)	126 (8)	p<0.05
Total colonoscopies	267 (15)	1509 (85)	NA
Normal colonoscopy	114 (43)	516 (34)	p<0.01
Diverticulosis	58 (22)	442 (29)	p<0.05
Colonic polyps	55 (21)	334 (22)	p=0.58
Colonic inflammation	29 (11)	230 (15)	p=0.06

NA = not applicable.

Therefore, while these methods may be scientifically pure, they are not practical to achieve a broader picture of endoscopy reporting in normal clinical practice.

Gastroscopy reporting has been studied more extensively than colonoscopy reporting. In particular, the degree of oesophagitis has been investigated by several authors.^{11–13,16,18} In published studies, kappa values are often below 0.4, signifying poor inter-observer agreement. This is often the case with inexperienced endoscopists,^{11,12} and for more minor degrees of oesophagitis.¹³ Agreement on the complications of oesophagitis (stricture and ulceration) is better.^{11,13} In addition, intraobserver agreement is often poor,¹² so perhaps it is not surprising that different endoscopists cannot agree. Our own study showing large variations in the frequency of reporting oesophagitis (all grades) would be consistent with previous studies.

Data are more sparse for other upper gastrointestinal endoscopic diagnoses: studies have shown conflicting results in the case of oesophageal varices,^{14,20} while gastric varices are inconsistently reported.²⁰ Experienced endoscopists are able to agree about the presence of deep gastric ulceration in patients not taking non-steroidal anti-inflammatory drugs (NSAIDs), but agreement was poor on the presence of ulceration in those taking NSAIDs.²¹ Expert endoscopists are good at agreeing on the presence of active ulcer bleeding and the stigmata of recent bleeding.¹⁹ Trainees, however, are less good.²² The present study highlights significant differences in the reporting of peptic ulceration of all aetiologies.

Recent studies of colonoscopy reporting have focused on completion rates rather than interpretation of endoscopic findings.^{15,23,24} When experienced endoscopists review photographs purporting to show caecal landmarks, there is considerable disagreement as to whether the photographs document total colonoscopy.²⁴ In all the studies, caecal intubation rates differed widely between endoscopists as they did in the present study. Completion rates improve as endoscopic experience increases.²³

Furthermore, some of the variance may be explained by the endoscopists' perceived need to reach the caecum.²⁵

The largest study of colonoscopy reporting also found that the frequency of normal examinations varied from less than 10% to greater than 40%.¹⁵ The authors speculate that this may have been due to endoscopists having different thresholds and indications for doing a colonoscopy. Our study used a joint waiting list, yet the frequency of normal examinations for colonoscopy varied from 23–55%. This would suggest that most of the difference in the rate of reporting normal examinations is likely to be due to different interpretation of the endoscopic findings, rather than different patient selection. Likewise, the variance in the recording of normal gastroscopies from 23–59% is also likely to

be due to differing interpretation, rather than different patient groups.

Polyp detection in patients undergoing surveillance colonoscopy has also been reported to vary significantly between endoscopists.¹⁵ Our own endoscopists found polyps in 14–45% of cases (all indications). The variation in polyp reporting is of concern as it may have serious implications for the patients' future health. Accurate polyp detection is vital if national colorectal cancer screening is to be effective.

Two studies have examined variable reporting of colonic inflammation.^{17,26} In Crohn's disease there is good agreement on the presence of ileal ulceration and stricturing as well as on most of the endoscopic features of Crohn's colitis.²⁶ For ulcerative colitis,¹⁷ agreement is generally good, but poor for mild to moderate activity and stricturing. Our own study found that inflammation overall (as opposed to specific inflammatory features) was very inconsistently recorded.

Previous studies have suggested that inexperienced endoscopists show greater interobserver variability.^{16,17} In our study, trainees were more likely to report a normal examination, and reported peptic ulceration, oesophagitis and diverticulosis, but not colonic polyps, less frequently than specialists. This may be due to inexperience in detecting lesions that are subtle or difficult to spot. This highlights the importance of adequate training and supervision.

A potential weakness of our study is that the casemix of the different endoscopists may have varied, thus explaining the variation in final diagnosis reporting. Since the endoscopies were all listed from a joint waiting list, differences should not have been great. It remains possible that there may have been a degree of selection bias on individual lists, due to endoscopists' special interests, for example, but it is unlikely to have made a significant difference to the overall results. The strength of this study is the large numbers of procedures analysed and highly significant differences found between specialists.

The data for the study were derived from the final endoscopy report so, with the exception of self-reported colonoscopy completion where there is a potential for bias, it is unlikely that the diagnosis reported would be other than that felt to be correct by the endoscopist. The overall colonoscopy completion rate of 82% is similar to the figure of 77% from a recent national prospective audit of colonoscopy practice,²⁷ suggesting that our institution is probably representative of the national situation. It is not possible to determine from the databases whether some conditions (eg hiatus hernias or haemorrhoids) were under-reported by certain endoscopists since they considered the findings trivial. While the clinical importance of some of the reported diagnoses is debatable, such inconsistency is in itself important, since GP and, in particular, patient perception of these conditions may be very different from that of the hospital specialist.

Our study does not address the potential influence that a computerised reporting system might have on the characterisation of lesions and the wording of the endoscopy report. The endoscopist is constrained by the reporting software and endoscopists might report findings with less variability in a narrative report. Computerised reporting allows reports to be archived and databases to be interrogated for audit purposes, but its influence on the recording and interpretation of individual reports deserves consideration.

The results of the present study highlight the variability in reporting common endoscopic diagnoses in normal clinical practice. This may have potentially important clinical consequences. To date, expert opinion has largely centred on technical competence and safe endoscopy. There is now a pressing need to develop and refine diagnostic standards. Fundamental to this should be an increased emphasis on training in endoscopic interpretation. Significant advances in the optical resolution of video endoscopes, and in the capture, storage, and reproduction of video endoscopic images in digital form are currently being introduced. We believe that these technical developments should be used for the development of standardised educational material, which should enhance training and ultimately improve patient care.

References

- Department of Health. Hospital Episode Statistics (HES) online. www.hesonline.nhs.uk
- Bramble MG, Suvakovic Z, Hungin AP. Detection of upper gastrointestinal cancer in patients taking antisecretory therapy prior to gastroscopy. *Gut* 2000;46:464–7.
- Postic G, Lewin D, Bickerstaff C, Wallace MB. Colonoscopic miss rates determined by direct comparison of colonoscopy with colon resection specimens. *Am J Gastroenterol* 2002;97:3182–5.
- Laine L, Cohen H, Sloane R, Marin-Sorensen M, Weinstein WM. Interobserver agreement and predictive value of endoscopic findings for H. pylori and gastritis in normal volunteers. *Gastrointest Endosc* 1995;42:420–3.
- American Society of Gastrointestinal Endoscopy (ASGE). Quality and outcome assessment in gastrointestinal endoscopy. *Gastrointest Endosc* 2000;54:662–7.
- O'Mahony S, Naylor G, Axon A. Quality assurance in gastrointestinal endoscopy. *Endoscopy* 2000;32:483–8.
- Marshall JB. Technical proficiency of trainees performing colonoscopy: a learning curve. *Gastrointest Endosc* 1995;42:287–91.
- Tassios PS, Ladas SD, Grammenos I, Demertzis K, Raptis SA. Acquisition of competence in colonoscopy: the learning curve of trainees. *Endoscopy* 1999;31:702–6.
- Endoscopy Training Guidelines. *Gastroenterology* 1996;110:1289–93.
- Joint Advisory Group on Gastrointestinal Endoscopy. www.thejag.org.uk
- Armstrong D, Bennett JR, Blum AL *et al.* The endoscopic assessment of oesophagitis: a progress report on observer agreement. *Gastroenterology* 1996;111:85–92.
- Kusano M, Ino K, Yamada T *et al.* Interobserver and intraobserver variation in endoscopic assessment of GERD using the 'Los Angeles' classification. *Gastrointest Endosc* 1999;49:700–4.
- Bytzer P, Havelund T, Hansen JM. Interobserver variation in the endoscopic diagnosis of reflux oesophagitis. *Scand J Gastroenterol* 1993;28:119–25.
- Bendtsen F, Skovgaard LT, Sorensen TI, Matzen P. Agreement among multiple observers on endoscopic diagnosis of esophageal varices before bleeding. *Hepatology* 1990;11:341–7.
- Cotton PB, Connor P, McGee D *et al.* Colonoscopy: practice variation among 69 hospital-based endoscopists. *Gastrointest Endosc* 2003;57:352–7.
- Pandolfino JE, Vakil NB, Kahrilas PJ. Comparison of inter- and intraobserver consistency for grading of esophagitis by expert and trainee endoscopists. *Gastrointest Endosc* 2002;56:639–43.
- Orlandi F, Brunelli E, Feliciangeli G *et al.* Observer agreement in endoscopic assessment of ulcerative colitis. *Ital J Gastroenterol Hepatol* 1998; 30:539–41.
- Lundell LR, Dent J, Bennett JR *et al.* Endoscopic assessment of oesophagitis: clinical and functional correlates and further validation of the Los Angeles classification. *Gut* 1999;45:172–80.
- Mondardini A, Barletti C, Rocca G *et al.* Non-variceal upper gastrointestinal bleeding and Forrest's classification: diagnostic agreement between endoscopists from the same area. *Endoscopy* 1998; 30:508–12.
- Cales P, Buscaill L, Bretagne JF *et al.* Interobserver and intercentre agreement of gastro-esophageal endoscopic signs in cirrhosis. Results of a prospective multicenter study. *Gastroenterol Clin Biol* 1989; 13:967–73.
- Hudson N, Everitt S, Hawkey CJ. Interobserver variation in assessment of gastroduodenal lesions associated with non-steroidal anti-inflammatory drugs. *Gut* 1994;35:1030–2.
- Bour B, Person B, Cales P *et al.* Interobserver agreement on endoscopic diagnosis of bleeding peptic ulcers. *Gastrointest Endosc* 1997;46:27–32.
- Dafnis G, Granath F, Pahlman L *et al.* The impact of endoscopists' experience and learning curves and interendoscopist variation on colonoscopy completion rates. *Endoscopy* 2001;33:511–7.
- Marshall JB, Brown DN. Photodocumentation of total colonoscopy: how successful are endoscopists? Do reviewers agree? *Gastrointest Endosc* 1996;44:243–8.
- Robinson J, Small P, Bell GD *et al.* A prospective audit of colonoscopy in a large DGH – factors affecting caecal intubation rates [abstract]. *Gut* 2001;48:A10.
- Smedh K, Olaison G, Jonsson KA *et al.* Interobserver variation of colono-ileoscopy findings in Crohn's disease. *Scand J Gastroenterol* 1995;30:81–6.
- Bowles CJ, Leicester R, Romaya C *et al.* A prospective study of colonoscopy practice in the UK today: are we adequately prepared for national colorectal cancer screening tomorrow? *Gut* 2004;53:277–83.