Histopathology 2009, **54**, 699–712. DOI: 10.1111/j.1365-2559.2009.03288.x

Barrett's dysplasia and the Vienna classification: reproducibility, prediction of progression and impact of consensus reporting and p53 immunohistochemistry

Philip V Kaye,[1,2] Syeda A Haider,[1] Mohammed Ilyas,[1,2] Peter D James,[1] Irshad Soomro,[1] Wasek Faisal,[2] James Catton,[2] Simon L Parsons[2] & Krish Ragunath[2]

Departments of [1]Cellular Pathology, and [2]Wolfson Digestive Diseases Centre, Nottingham University Hospitals and NHS Trust and University of Nottingham, Queens Medical Centre Campus, Nottingham, UK

Date of submission 4 July 2008
Accepted for publication 5 November 2008

Kaye P V, Haider S A, Ilyas M, James P D, Soomro I, Faisal W, Catton J, Parsons S L & Ragunath K (2009) *Histopathology* **54**, 699–712

Barrett's dysplasia and the Vienna classification: reproducibility, prediction of progression and impact of consensus reporting and p53 immunohistochemistry

Aims: The Vienna classification is used to classify dysplasia in Barrett's oesophagus (BO), but reproducibility and value of diagnosis of lower grades in particular are often questioned. The aim was to test the diagnostic variability and correlation with patient outcome and to attempt to define histological features causing discrepant diagnoses, as well as to test the impact of adding p53 immunohistochemistry on reproducibility and prediction of outcome.

Methods and results: One hundred and forty-three patients with 154 sets of biopsy specimens originally diagnosed with Barrett's dysplasia were retrieved from the pathology records of Nottingham University Hospital. Thirty-two Barrett's patients without dysplasia were added. Anonymized slides were graded independently by five pathologists without and with p53-stained slides. Interobserver variation, correlation with outcome and diagnostic accuracy were determined. Weighted κ scores between pairs of pathologists showed substantial agreement and improved after p53 immunohistochemistry. Agreement with the original diagnosis was substantially lower. Fourteen of 34 low-grade dysplasias (LGD) and 27 of 30 high-grade dysplasias on consensus progressed within 10 years compared with 18/94 and 28/39 of original diagnoses. Progression correlated with p53 positivity.

Conclusion: The Vienna classification is useful and reproducible in BO. Consensus diagnosis by gastrointestinal pathologists produces high specificity and predictive value, even for LGD. p53 immunohistochemistry assists in diagnosis in difficult cases and predicts progression.

Keywords: Barrett, carcinoma, dysplasia, oesophagus, p53

Abbreviations: BO, Barrett's oesophagus; EMR, endoscopic mucosal resection; GI, gastrointestinal; H&E, haematoxylin and eosin; HGD, high-grade dysplasia; HR, hazard ratio; ID, indefinite for dysplasia; LGD, low-grade dysplasia; ND, no dysplasia; NPV, negative predictive value; PPV, positive predictive value; RR, relative risk

Introduction

Barrett's oesophagus (BO) is a premalignant condition largely responsible for the increase in oesophageal adenocarcinoma in Europe and North America. Progression is believed to be a multistep process, which progresses often over many years through increasing degrees of epithelial atypia.[1] Intraepithelial neoplasia is recognized histologically as dysplasia, but this is complicated by the difficulty in diagnosing lesser degrees of dysplasia as features may overlap with non-neoplastic regenerative changes,[2] and some studies have indicated that this intraepithelial neoplasia

Address for correspondence: Dr Philip Victor Kaye, Department of Cellular Pathology, Nottingham University Hospitals, Queens Medical Centre Campus, Derby Rd, Nottingham, NG7 2UH, UK.
e-mail: philip.kaye@nuh.nhs.uk

may remain static or even regress.[3] Nevertheless, there is currently no other method for diagnosis of intraepithelial neoplasia or predicting progression to carcinoma. Patients with BO are generally entered into endoscopic surveillance programmes and, despite acknowledged difficulties in diagnosis, follow-up and treatment are largely based on histological reporting of random biopsy specimens.[4]

The revised Vienna classification was introduced as a way of ensuring global uniformity of reporting of epithelial changes in gastrointestinal (GI) dysplasia, including BO.[5] It is largely based on Riddell's classification[6] introduced in North America in the 1980s, with some minor changes (Table 1). Although there have been few studies specifically validating the Vienna classification in BO, there are several studies based on the Riddell schema and these are directly comparable.[7–9] These studies have shown variable reproducibility due to interobserver disagreement, particularly at the lower end of the spectrum, and many clinicians regard the diagnosis of low-grade dysplasia (LGD) with scepticism, believing it to be poorly predictive of progression.

Several biomarkers have been used in attempts to strengthen the diagnosis of dysplasia. These include DNA ploidy by flow cytometry,[10] chromosomal aberrations,[11] and p53[12] and α-methylacyl coenzyme A racemase immunohistochemistry.[13] p53 immunohistochemistry has been the most widely used and of the generally available tools provides the best evidence thus far for correlation with outcome.

With this in mind, our own cases of Barrett's dysplasia were analysed with the aim of testing interobserver variability and correlation with outcome. The impact of p53 immunohistochemistry on interobserver variability and on prediction of outcome was also examined. Reasons for discrepant diagnoses between pathologists were analysed in order to better define histological features that correlated (or did not correlate) with progression.

Table 1. Revised Vienna classification

Category	Diagnosis
1	Negative for neoplasia (ND)
2	Indefinite for neoplasia (ID)
3	Mucosal low-grade neoplasia (LGD) Low-grade adenoma Low-grade dysplasia
4	Mucosal high-grade neoplasia (HGD)
4.1	High-grade adenoma/dysplasia
4.2	Non-invasive carcinoma (carcinoma *in situ*)
4.3	Suspicious for invasive carcinoma
4.4	Intramucosal carcinoma
5	Submucosal invasion by carcinoma

Methods

CASE SELECTION

The Pathology Databases at Queens Medical Centre and City Hospital, Nottingham were searched between 1987 and 2004 for all cases diagnosed as oesophageal glandular dysplasia or possible dysplasia. Cases with definite invasive adenocarcinoma were excluded. Thirty-two random cases of BO without dysplasia or progression on at least 10 years' follow-up were also included.

HISTOLOGY AND p53

Haematoxylin and eosin (H&E)-stained slides were extracted from the files and one pathologist (S.A.H.) selected the most representative (atypical) one or occasionally two slides. These were anonymized and randomly numbered. Representative blocks were retrieved and stained for p53 with D07 antibody (Dako, Ely, UK) using the Dako Techmate System (Dako) with microwave antigen retrieval. Pathologists included four specialist GI/oesophageal pathologists and one senior trainee pathologist. Slides were circulated to each pathologist in the study (M.I., S.A.H., P.D.J., I.S. and P.V.K.), who classified each case using the revised Vienna Classification (Table 1). Each case was then rescored after examining p53-stained slides. Pathologists were not asked to score the p53 itself, but to integrate the information it yielded into a final category with the H&E. One participant (P.V.K.) scored each p53-stained section as positive, negative or not representative. This was done qualitatively, paying particular attention to intensity of p53 immunoreactivity in areas of histological abnormality compared with remaining histologically non-dysplastic glands, which served as internal controls.

After all scoring had been completed, a consensus score for each case was determined. This was defined as the closest category to the average of the post p53 score of each pathologist.

CLINICAL OUTCOME

Patient notes or computerized records were examined to determine length of follow-up and further histological progression. The main clinical outcomes recorded were a definitive procedure for BO indicated by cancer or high-grade dysplasia (HGD) (usually oesophagectomy,

but also endoscopic mucosal resection (EMR) and radiotherapy in a few cases) or death from BO without a procedure (in advanced cases where palliative care only was offered and where a definite diagnosis of invasive cancer had been subsequently made).

STATISTICAL ANALYSIS

This was done using the SYSTAT programme (SAS Institute Inc., Cary, NC, USA). Diagnostic variability was analysed using Cohen's κ score in two ways: first, a multirater general agreement and category by category yielding unweighted κ scores, second, a weighted κ score between pairs of pathologists over all four categories. Pairwise comparison was also made with the consensus diagnosis and the original diagnosis in the pathology report. In this part of the study, some of the cases were biopsy specimens from the same patient taken at different times.

Correlation with outcome was restricted to the first specimen originally diagnosed as dysplastic. Patients lost to follow-up or deceased from causes unrelated to BO were treated as a negative outcome for the period in follow-up, as were patients alive and who had not had a procedure (including some with persistent dysplasia on most recent biopsy). Logrank (peto) test was used to determine relative risk (RR) of each category for progression over time versus whole series and hazard ratios (HR) for each category against baseline (no dysplasia). For sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) the outcome at 10 years after biopsy was used, restricted to those patients for whom these follow-up data were available. Figures were calculated for each pathologist, consensus and original diagnosis for each category against the baseline [no dysplasia (ND)].

CASE REVIEW

Cases that showed the most major disagreements on blinded analysis between pathologists were subsequently reviewed by participating pathologists to analyse reasons for differences. The main areas of disagreement were evaluated in the knowledge of clinical outcome, in order to identify features that may or may nor predict progression.

Results

DIAGNOSTIC VARIABILITY

One hundred and sixty-three sets of biopsy specimens with dysplasia were initially retrieved from the pathology databases together with 32 randomly selected cases without dysplasia. Nine were excluded, mainly because they were not BO but rather squamous dysplasia or gastric dysplasia. This left 175 patients with 186 sets of specimens for analysis in the diagnostic variability study. The multirater unweighted κ scores before and after p53 are shown in Table 2, by category and overall. There was moderate overall agreement, which improved with p53 immunohistochemistry. When consultants only were analysed, the overall agreement also improved. As expected, when analysed by category there was moderate agreement for categories 1 (ND) and 4 (HGD) and fair agreement for category 3 (LGD). Category 2 [indefinite for dysplasia (ID)] yielded only slight agreement. Weighted κ values are generally believed to give a better indication of agreement than unweighted values, because they penalise smaller degrees of disagreement less. These are shown on a pairwise basis in Table 3, before and after p53. All κ values improved with p53 immunohistochemistry. There was substantial agreement between pairs of pathologists and substantial to near-perfect agreement between pathologists and the consensus diagnosis. Agreement between pathologists and original reported diagnoses was considerably worse.

Table 2. Multirater unweighted κ, overall and by category without and with (bold) p53

	All		Consultants	
Category	No p53	p53	No p53	p53
1	0.56	**0.63**	0.58	**0.66**
2	0.08	**0.08**	0.1	**0.06**
3	0.23	**0.31**	0.26	**0.37**
4	0.65	**0.63**	0.66	**0.7**
Combined	0.42	**0.48**	0.45	**0.52**

CORRELATION WITH OUTCOME

Of 175 patients included in the survival analysis, 32 were never diagnosed with dysplasia and were known to have not developed Barrett's cancer. Fifty-one patients progressed to BO-related definitive treatment or death. Of these 51 patients, 31 progressed within 1 year of this diagnosis of dysplasia (seven of these had previous surveillance biopsies over a period longer than 1 year) and 20 were in a follow-up surveillance programme for between 12 and 216 months after this diagnosis of dysplasia (incident cases of cancer). The

Table 3. Weighted κ between pairs of pathologists without and with (bold) p53

	1		2		3		4		5		Original		Consensus	
1	–	–	0.50	**0.55**	0.62	**0.68**	0.55	**0.59**	0.59	**0.70**	0.30	**0.34**	0.67	**0.69**
2	0.50	**0.55**	–		0.52	**0.58**	0.53	**0.53**	0.58	**0.65**	0.31	**0.24**	0.60	**0.69**
3	0.62	**0.68**	0.52	**0.58**	–	–	0.58	**0.62**	0.62	**0.69**	0.28	**0.31**	0.70	**0.80**
4	0.55	**0.59**	0.53	**0.53**	0.58	**0.62**	–	–	0.65	**0.67**	0.34	**0.33**	0.76	**0.77**
5	0.59	**0.70**	0.58	**0.65**	0.62	**0.69**	0.65	**0.67**	–	–	0.36	**0.36**	0.8	**0.86**
Ori	0.30	**0.34**	0.31	**0.24**	0.28	**0.31**	0.34	**0.33**	0.36	**0.36**	–	–	0.35	**0.35**
Con	0.67	**0.69**	0.60	**0.69**	0.70	**0.80**	0.76	**0.77**	0.80	**0.86**	0.35	**0.35**	–	–

Ori, original diagnosis; Con, consensus diagnosis.

Table 4a. Original diagnosis on index biopsy with patient outcomes in those with Barrett's oesophagus (BO)-related procedure or death

Vienna	Surgical procedure: HGD	Surgical procedure: AC	No surgical procedure: AC (histo)	Procedure: LGD	No surgical procedure: AC (no histo)	
original						
1	0	0	0	0	0	0
2	0	0	0	0	1	1
3	4	10	4	2	2	22
4	5	18	3	0	2	28
Total	9	28	7	2	5	51

Table 4b. Consensus diagnosis and p53 immunohistochemistry on index biopsy with patient outcomes in those with BO-related procedure or death

Vienna Category	Surgical procedure: HGD	Surgical procedure: AC	No surgical procedure: AC (histo)	Procedure: LGD	No surgical procedure: AC (no histo)	Total
Consensus						
1	1	3	0	0	2	6
2	0	1	2	1	0	4
3	5	5	3	0	1	14
4	3	19	2	1	2	27
p53						
negative	1	9	2	0	3	16
positive	7	14	4	1	2	26

HGD, high-grade dysplasia; AC, adenocarcinoma; LGD, low-grade dysplasia.

outcomes in these 51 patients in relation to original and consensus index diagnosis are shown in more detail in Table 4a,b. The final given histological diagnosis (either at surgical procedure or last recorded biopsy) according to their original or consensus diagnosis is shown in Table 5a,b. The differences in proportions of LGD progressing between original and consensus diagnosis is noteworthy. The proportions of

Table 5a. Original diagnosis on index biopsy with reported histology on last follow-up biopsy or definitive procedure

	Last given histological diagnosis on follow-up or oesophagectomy/EMR						
Vienna category	ND	ID	LGD	HGD	AC	No further histology	Total
Original							
1	24	0	0	0	0	5	29
2	2	0	0	1	0	3	6
3	42	0	12	6	14	25	99
4	3	1	4	5	21	7	41
Total	71	1	16	12	35	40	175

Table 5b. Consensus diagnosis and p53 immunohistochemistry on index biopsy with reported histology on last follow-up biopsy or definitive procedure

	Last given histological diagnosis on follow-up or oesophagectomy/EMR						
Vienna category	ND	ID	LGD	HGD	AC	No further histology	Total
Consensus							
1	55	1	4	1	3	24	88
2	6	0	2	0	3	7	18
3	9	0	8	7	8	4	36
4	1	0	2	4	21	5	33
p53							
negative	54	0	5	2	11	27	99
positive	5	0	6	9	18	5	43

EMR, endoscopic mucosal resection; ND, no dysplasia; ID, indefinite for dysplasia; LGD, low-grade dysplasia; HGD, high-grade dysplasia; AC, adenocarcinoma.

patients progressing by 10 years for each category for each pathologist, consensus and original, is shown in Figure 1. Figure 2 shows progression by p53 status and number of pathologists diagnosing dysplasia.

The RR for progression over time for each pathologist by category (versus all cases) as well as for consensus diagnosis and original diagnosis is shown in Table 6 and the HRs for each category against baseline (ND) are shown in Table 7.

There are clear differences in the extent to which different pathologists' diagnoses correlated with progression, particularly for the ID and LGD categories. Furthermore, correlation with progression was much higher for the consensus diagnosis of LGD (RR 1.56, HR) than for the original diagnosis of LGD (RR 0.74). When incident cases only were analysed, this was even more striking, with a consensus diagnosis of LGD (RR 3.1) having much higher correlation than the original diagnosis of LGD (RR 1.07) and approaching the risk of progression of HGD (4.47).

The impact of the number of pathologists diagnosing dysplasia (LGD or HGD) was also evaluated (Tables 4 and 5). It is clear that where four of five (RR 1.9, HR 7.7) or all five pathologists (RR 3.64, HR 14) diagnosed dysplasia, progression was much more likely than when fewer pathologists did so. This was even more striking in incident cases.

Comparison of p53 immunohistochemistry with consensus Vienna category is shown in Table 8 (bearing in mind that the consensus diagnosis might have been influenced by the p53 immunohistochemistry). The outcomes and final histological diagnoses are shown in Tables 4b and 5b with reference to the initial p53 immunohistochemistry. The effect of p53 positivity is also shown in Tables 6 and 7 and confirms its predictive nature. In most cases that were positively scored there was a clear-cut dichotomy between strongly positive abnormal glands and faintly stained background epithelium, whereas negative cases were either completely negative or showed weak immunopositivity at the base of crypts in all glands. A few cases were equivocal, either because of strong diffuse background reactivity or weaker focal reactivity, and these were scored as negative. In some cases, the atypical area was stained immunopositive, with a few less atypical surrounding glands also positive. Unfortunately, in 33/175 cases representative material was not present in the p53-stained sections because the atypical area had cut out. Of note, although p53 strongly correlated with histological dysplasia and outcome, p53– dysplasia was not uncommon.

Another way to express the usefulness of the dysplasia diagnosis is to look at the specificity, sensitivity, PPV and NPV. This is shown in Table 9 with each category compared with ND. Again, it is clear that the consensus diagnosis of especially LGD has a much better PPV than the original diagnosis, and the impact of more pathologists agreeing is similar. It was not possible to compare NPV and sensitivity for the original diagnosis as only originally diagnosed positive cases together with a few negative cases that were known not to have progressed were included.

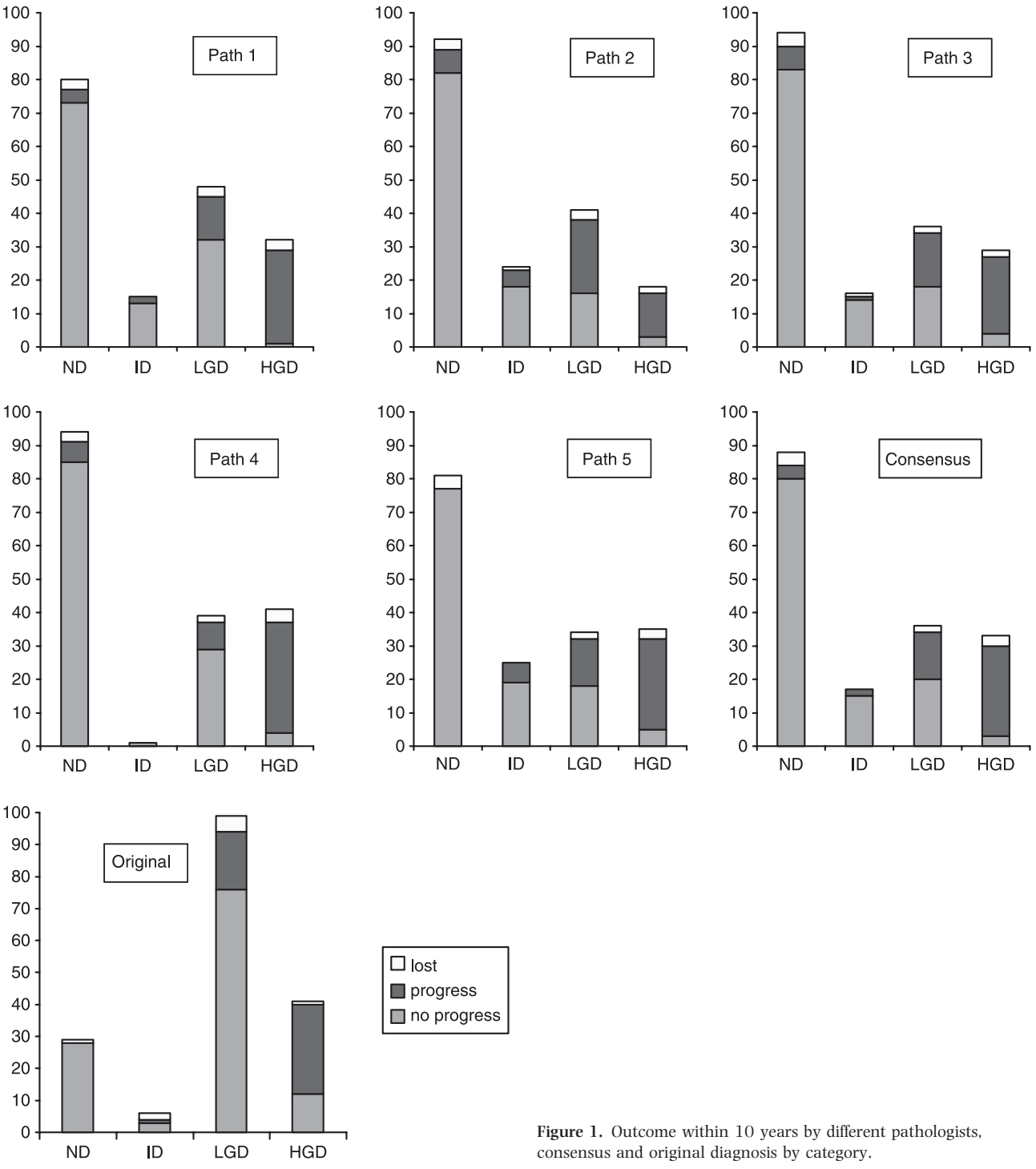

Figure 1. Outcome within 10 years by different pathologists, consensus and original diagnosis by category.

CASE REVIEW

Forty-seven cases showing substantial diagnostic variation were reviewed in the light of subsequent progression. The following features caused diagnostic difficulty: (i) the most common cause of diagnostic variation was in interpretation of multilayering in the surface epithelium (24 cases). In particular, pseudostratification by elongated pencil-like nuclei was interpreted by some as dysplasia, but this consistently did not

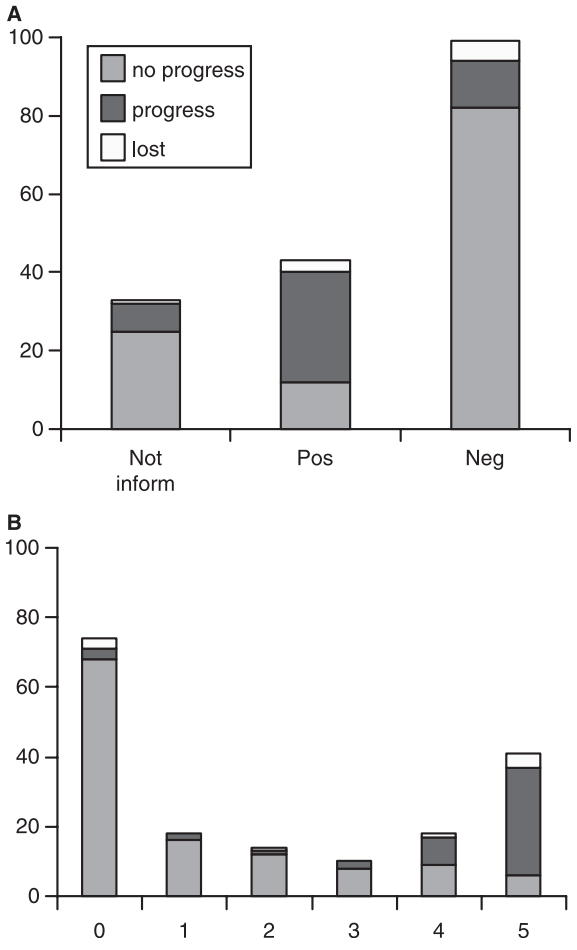

Figure 2. Outcome within 10 years by (**A**) p53 status and (**B**) number of pathologists diagnosing dysplasia.

correlate with progression (Figure 3), (ii) ulceration or marked inflammation. Some pathologists downgraded the dysplasia in these cases more than others, while there was variable attention paid to the degree of nuclear atypia. Some of these ulcerated/inflamed cases in fact did correlate with rapid progression, especially if severe nuclear atypia was present (Figure 4), (iii) focal areas of architectural or cytological atypia that stood out from the background epithelium were more likely to represent true dysplasia than similar changes present throughout the biopsy specimen (Figure 5), and (iv) apparent surface maturation. Some would not diagnose dysplasia without seeing surface involvement or in the presence of apparent maturation, whereas others regarded this as a relative marker. Certainly, many cases with undisputed dysplasia showed only patchy surface involvement, with other areas showing apparent maturation. Cases where the diagnosis of dysplasia was not in doubt were also reviewed to explore criteria for the distinction between LGD and HGD. Reasons for this variation were much more difficult to identify, probably because this division involves selecting an unnatural cut-off along a biological continuum. Variation was therefore due to different weightings given to degrees of cytological and architectural atypia. Of particular interest were dysplastic lesions with an exophytic villous morphology. Pathologists tended to classify these as HGD regardless of whether or not the nuclei showed high-grade features (Figure 6). In fact, this approach was borne out by the subsequent oesophagectomies, which often showed underlying invasive tumour. The criteria for

Table 6. Relative Risk for BO death/procedure for whole series and incident cases (bold)

	Cat. 1		Cat. 2		Cat. 3		Cat. 4	
Path 1	0.26	**0.24**	0.56	**0.77**	1.02	**1.76**	4.76	**7.54**
Path 2	0.35	**0.42**	0.78	**1.2**	2.2	**3.34**	4.03	**2.33**
Path 3	0.34	**0.45**	0.41	**1.04**	1.77	**3.5**	4.24	**2.83**
Path 4	0.31	**0.31**	0.39	**0**	0.79	**0.79**	3.94	**3.94**
Path 5	0.1	**0.20**	1.15	**1.56**	1.7	**3.55**	3.84	**3.90**
Ori	–	–	–	–	0.74	**1.07**	3.12	**3.06**
Con	0.225	**0.2**	0.75	**1.8**	1.56	**3.10**	4.1	**4.47**

	0		1		2		3		4		5	
n PDD	0.27	**0.41**	0.31	**0**	0.45	**1.1**	0.75	**1.16**	1.9	**3.95**	3.64	**4.6**
p53	0.83	**1.04**	3.1	**5.47**	0.48	**0.46**	–	–	–	–	–	–

Ori, original diagnosis; Con, consensus diagnosis; *n* PDD, number of pathologists diagnosing dysplasia; BO, Barrett's oesophagus.

Table 7. Hazard ratios for whole series and incident cases (bold)

	Cat. 2 versus 1		Cat. 3 versus 1		Cat. 4 versus 1	
Path 1	2.2 (0.86–5.6)	**3.2 (0.83–13)**	4 (2–7.5)	**7.25 (2.6–21)**	18.5 (7.5–45)	**31 (3.8–263)**
Path 2	2.2 (1–4.9)	**2.9 (0.9–9.4)**	6 (3–12)	**7.8 (2.3–28)**	11.4 (3.6–36)	**5.5 (0.26–116)**
Path 3	1.2 (0.47–3.2)	**2.3 (0.51–10.3)**	5.3 (2.5–11)	**7.8 (2.1–29)**	12.6 (5.1–31)	**6.2 (0.87–45)**
Path 4	–	–	2.6 (1.3–5)	**2.8 (0.96–8.1)**	12.5 (5.9–27)	**13 (2.6–59)**
Path 5	11.4 (5–26)	**7.6 (2–29)**	16.9 (7.8–37)	**17.36 (4.3–71)**	38.2 (16.8–86)	**19 (3.6–100)**
Con	3.3 (1.3–8.3)	**9 (2.2–50)**	7.1 (3.3–14)	**15.4 (4.1–59)**	20 (8.3–50)	**22 (3.8–129)**

	0 versus 1	0 versus 2	0 versus 3	0 versus 4	0 versus 5
n PDD	1.2 (0.5–2.8)	1.7 (0.62–4.7)	2.8 (0.79–10)	7.7 (2.6–20)	14 (6.5–31)
	–	**2.7 (0.57–14)**	**2.8 (0.32–25)**	**9.6 (1.3–77)**	**11 (2.5–53)**
p53	3.8 (1.5–10)	–	–	–	–
	5.2 (0.9–30)	–	–	–	–

Con, consensus diagnosis; *n* PDD, number of pathologists diagnosing dysplasia.

Table 8. p53 immunohistochemistry in relation to consensus Vienna category

	p53			
Vienna Category	Insufficient	Pos.	Neg.	Total
Consensus				
1	10	0	78	88
2	7	1	10	18
3	12	21	3	36
4	4	21	8	33
Total	33	43	99	175

diagnosing HGD in the oesophagus therefore seem different from those in the colon, where some of these lesions would be categorized as low-grade adenomas.

Discussion

Despite many innovations in endoscopy and molecular techniques, pathology is still regarded as the gold standard for defining risk of progression in BO. The Vienna classification was introduced to ensure uniform terminology globally in the diagnosis of dysplasia. Two important studies using the similar Riddell classification, one focusing on interobserver variation[8] and the other on outcome[14], did show reasonable agreement between pathologists, particularly in differentiating HGD from lower grade lesions and in predicting prognosis. Lower grade lesions, particularly the interface between LGD and regenerative changes, remain a challenging area and typically show poor or slight agreement. A small but important study[15] has suggested that obtaining agreement on the presence of dysplasia by multiple pathologists as well as p53 positivity by immunohistochemistry might predict likelihood of progression in LGD.

This study has shown good agreement between pairs of pathologists with weighted κ values over all categories, varying between 0.53 and 0.70. Kappa values improved with use of p53 immunohistochemistry, suggesting that using this routinely could improve agreement. Weighted κ scores between pairs of reviewing pathologists were much better than between reviewing pathologists and the original diagnoses. This suggests that there has been a tightening of the diagnostic criteria over the years, possibly due to a shift to subspecialist reporting. Multirater unweighted κ scores were reasonable and also improved after taking p53 immunohistochemistry into account, and were also better when only consultants (all subspecialist) were included. It is difficult to compare κ scores between different studies, as these are influenced by the distribution of cases in the study as well as whether weighted or unweighted κ values were used. However, the most comparable study to this one was that of Montgomery *et al.*,[8] utilising 12 expert GI pathologists.

Table 9. Sensitivity, specificity, NPV and PPV for outcome within 10 years of diagnosis of dysplasia by category

	Outcome	Sensitivity	Specificity	PPV	NPV
Path 1 ID = 15	2	33	85	13	95
Path 2 ID = 23	5	42	82	22	92
Path 3 ID = 15	1	12.5	86	7	92
Path 5 ID = 25	6	100	80	24	100
Ori ID = 4	1	–	90	25	–
Cons ID *n* = 17	2	33	84	12	95
Path 1 LGD *n* = 45	13	76	70	29	95
Path 2 LGD *n* = 38	22	76	84	58	92
Path 3 LGD *n* = 34	16	70	82	47	92
Path 4 LGD *n* = 37	8	57	75	22	93
Path 5 LGD *n* = 32	14	100	81	44	100
Ori LGD *n* = 94	18	–	27	19	–
Cons LGD *n* = 34	14	78	80	42	95
Path 1 HGD *n* = 28	27	88	99	97	95
Path 2 HGD *n* = 16	13	65	96	81	92
Path 3 HGD *n* = 27	23	77	95	85	92
Path 4 HGD *n* = 37	33	85	96	89	93
Path 5 HGD *n* = 32	27	100	94	84	100
Ori HGD *n* = 39	28	–	70	70	–
Cons HGD *n* = 30	27	87	96	90	95
p53 pos n *n* = 40	28	80	68	70	78
3/5 dysplasia *n* = 10	2	40	89	20	96
4/5 dysplasia *n* = 17	8	72	88	47	96
5/5 dysplasia *n* = 37	31	91	92	84	96

PPV, positive predictive value; NPV, negative predictive value; ID, indefinite for dysplasia; Ori, original diagnosis; Con, consensus diagnosis; LGD, low-grade dysplasia; HGD, high-grade dysplasia.

This showed a κ value of 0.43 for the four diagnostic categories used in this study. This is similar to the unweighted κ value of 0.45 which we found amongst our GI consultants without the benefit of p53 immunohistochemistry. Some studies have reported substantially poorer agreement[9]; the reasons for this are uncertain. The correlation with progression is of great importance in advising clinicians on the risks associated with a particular category. Our data showed that the degree with which each individual pathologist was able to predict outcome (oesophagectomy or Barrett's carcinoma-related death) varied considerably, especially in LGD and ID. However, all produced better specificity for LGD than the original diagnoses, and the consensus diagnoses were much more specific and predictive. Again, this suggests improvement over time in the diagnosis of even the lower grades of dysplasia and underlines the value of consensus diagnosis in this

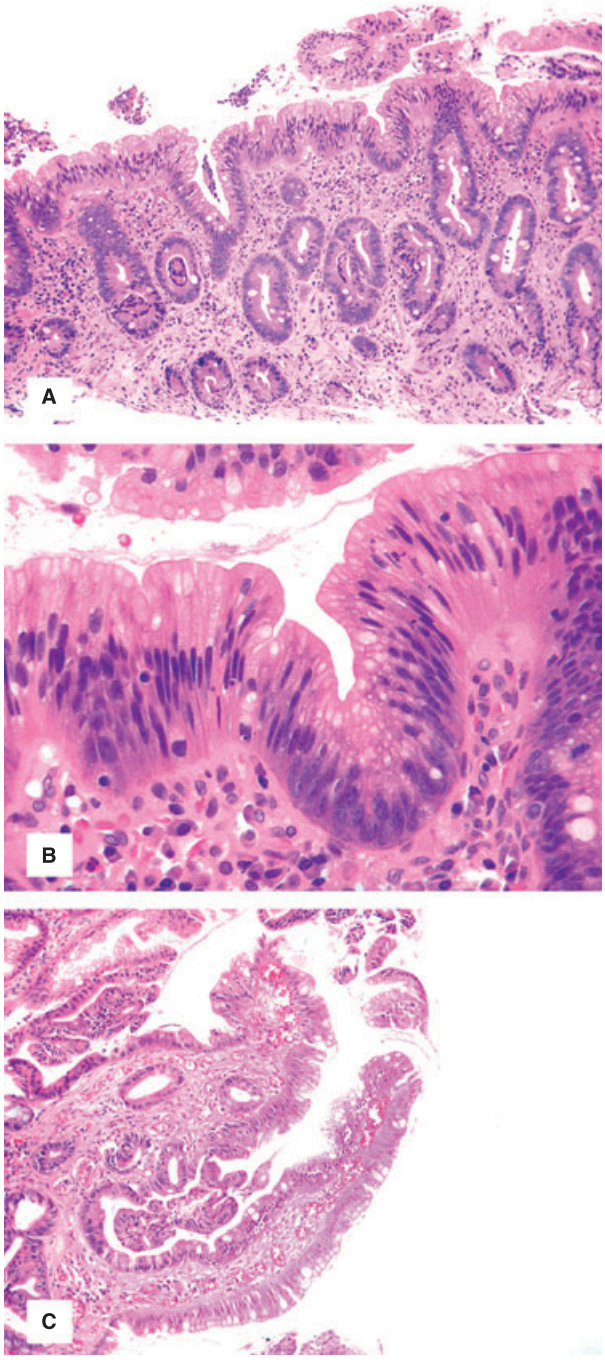

Figure 3. **A**,**B**, Pseudostratification with pencil-like nuclei – this should not be diagnosed as dysplasia. **C**, Villous configuration of surface with pseudostratified nuclei – this is a regenerative feature.

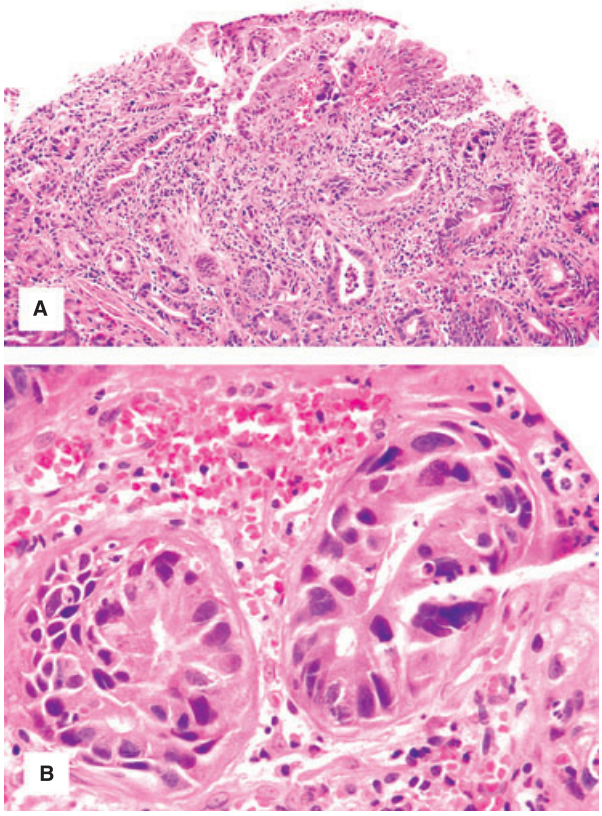

Figure 4. **A**,**B**, Heavy inflammation with high-grade atypia. Severe nuclear atypia such as depicted here should not necessarily be downgraded because of inflammation.

area. Fourteen of 34 (41%) cases of LGD on consensus progressed to oesophagectomy or Barrett's carcinoma-related death within 10 years compared with a 18/94 (19%) of original LGD diagnoses. For HGD, the consensus diagnosis was also more likely to be associated with progression (27/30, 90%) than the original diagnosis (28/39, 72%).

It was also striking that progression was closely related to the number of pathologists agreeing a diagnosis of dysplasia (high or low grade). There was often discordance in the grade of dysplasia between pathologists, but when all five pathologists agreed that dysplasia of any grade was present there was a very high rate of progression (84%), and this was 47% if 4/5 pathologists concurred. Lesser agreement was much less likely to progress.

There is some evidence to support the use of p53 immunohistochemistry as an adjunct for diagnosing dysplasia.[15–17] In this study, where material was available, p53 immunohistochemistry improved inter-observer variation in dysplasia diagnosis and correlated strongly with progression. However, p53–dysplasia is not uncommon, and negativity should not deter dysplasia diagnosis in a histologically

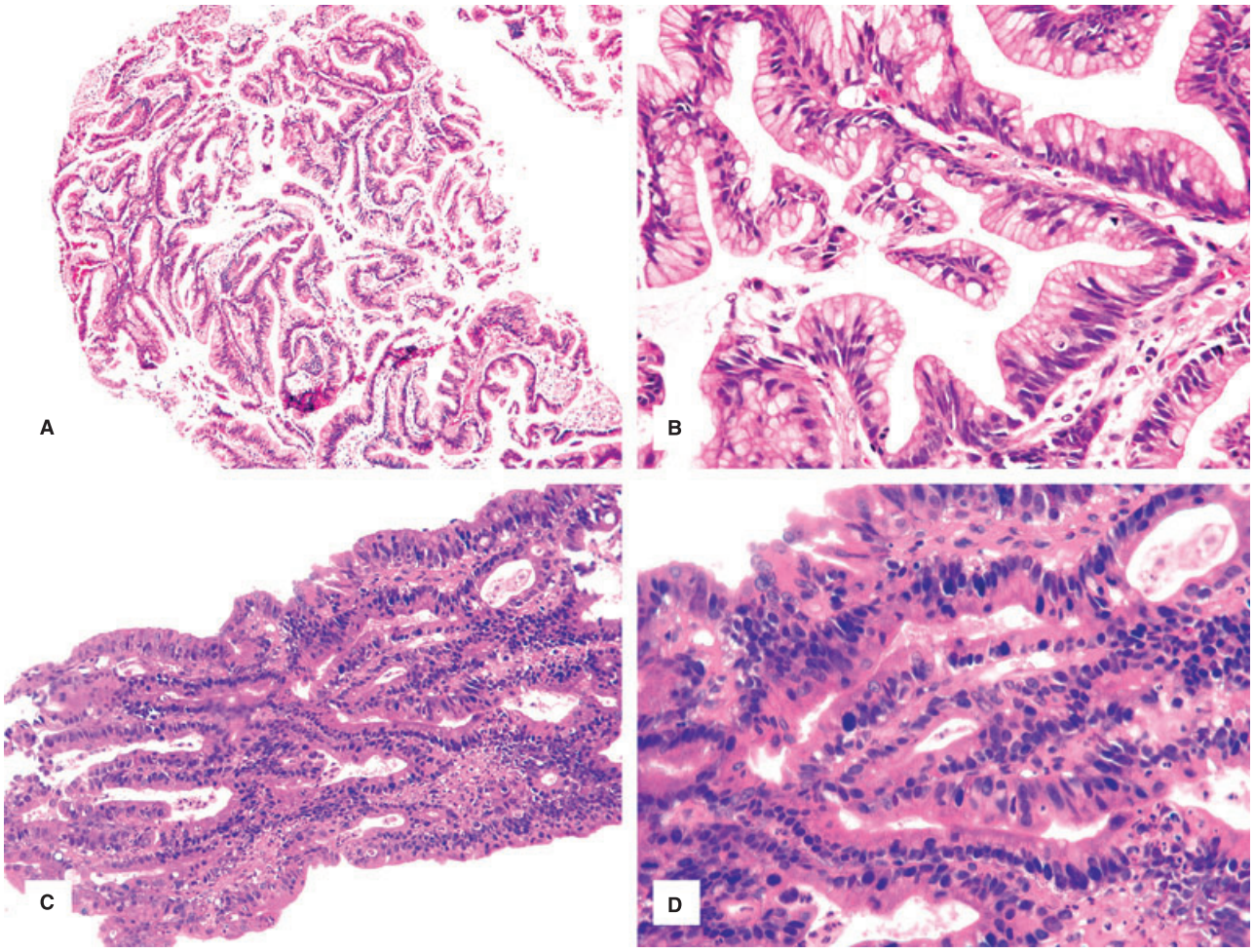

Figure 5. This shows two cases (**A**,**B**) and (**C**,**D**) with stratification but with definite nuclear atypia. Focality is also present in both cases. Consensus diagnosis was low-grade dysplasia on both.

unequivocal case. In less definite cases, p53 positivity in the area of interest does correlate with likelihood of progression. In our cases of consensus LGD, 3/12 p53– cases progressed versus 11/21 p53+ cases. Some studies have used complicated scoring systems for judging p53 positivity.[18] In our experience this is not necessary, as in positive cases there was almost always a clear-cut difference between immunopositivity in areas of suspicion compared with negative or very weak positivity in background glands that act as an internal control. Occasional equivocal cases are best regarded as negative. As the intensity of p53 immunoreactivity is likely to vary between laboratories, semiquantitative scoring systems are unlikely to be easily translatable into clinical practice. However, if attention is paid to differentiating between reactivity in areas of interest relative to background non-dysplastic glands, where clear differences exist this is more likely to be significant and reproducible. In some cases, immunopositivity was seen not only in areas of histological dysplasia, but in a few surrounding glands as well. This suggests that changes in p53 expression or mutation may sometimes antedate histological dysplasia; this is in accordance with some previous work[17]. p53– dysplasia is not uncommon, and probably reflects those cases which either do not harbour mutant p53 or where the mutation results in loss of p53 expression, rather than the more usual protein stabilization. Likewise, p53 positivity on immunohistochemistry may not always indicate mutation, but in some cases could represent wild-type overexpression.[17] However, this probably still represents mucosa that is more unstable and likely to progress. In any event, these data suggest that p53 immunohistochemistry is a useful predictive marker and adjunct in dysplasia diagnosis.

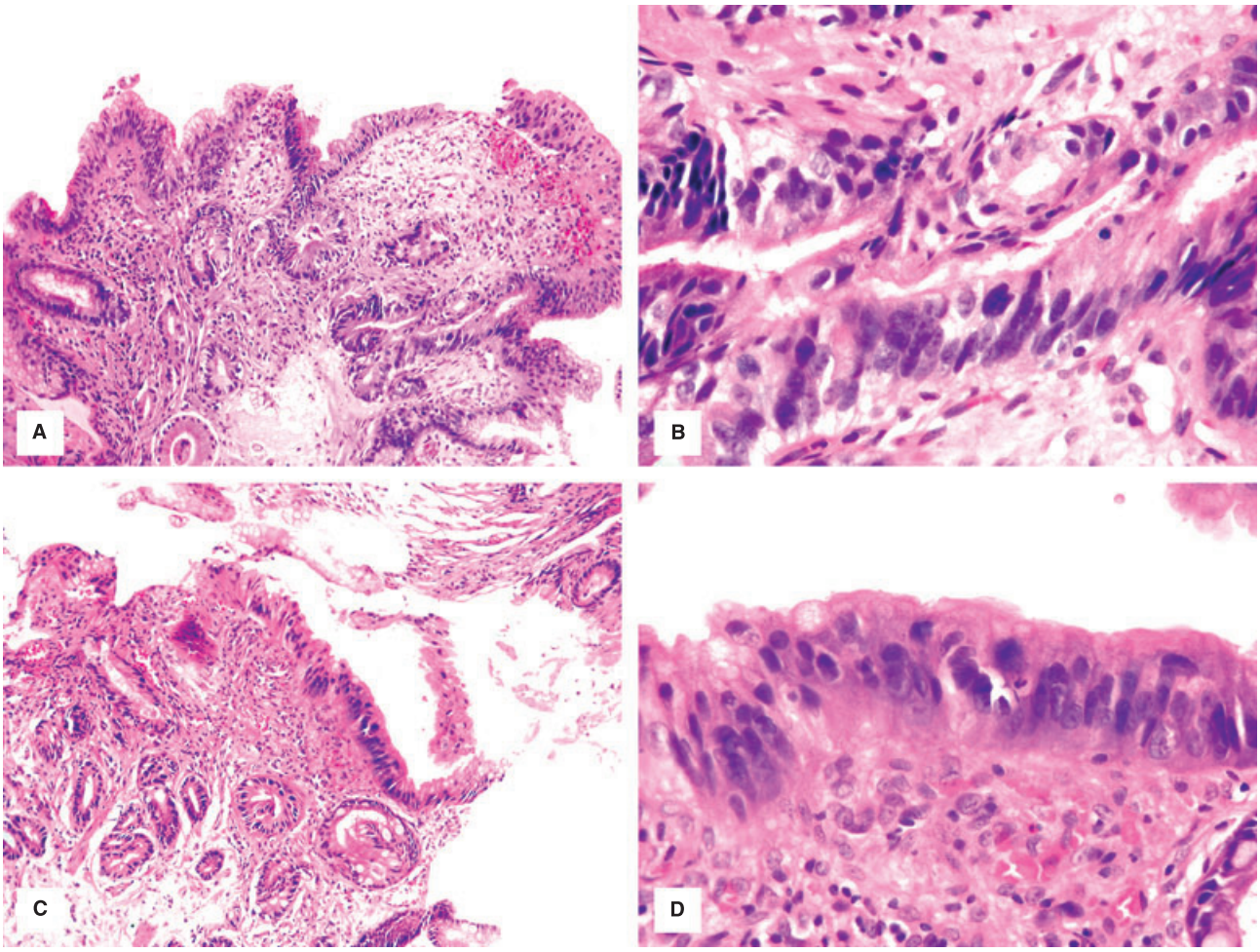

Figure 6. Two dysplastic exophytic lesions (**A**,**B**). This case was diagnosed variably as low-grade or high-grade dysplasia (HGD). **C**, **D**, This case was diagnosed as HGD by all pathologists. Both cases would probably be diagnosed as low-grade adenomas of the colon, but both showed advanced tumours at oesophagectomy.

As in most studies, the greatest variation in diagnosis came in category 2 (ID) with major variation in the frequency of use between pathologists and its consequent significance. One pathologist used this diagnosis 25 times, and six of these cases (24%) eventually required oesophagectomy whereas none of the NDs did. Another pathologist did not use this category at all. The latter strategy tended to result in relative over-diagnosis of LGD, which then became less predictive of progression while also allocating some cases with atypia, possibly indicating an underlying lesion, to ND. It is therefore important to use this category for any case where there is diagnostic doubt, rather than feeling obliged to make an apparently more decisive (but possibly incorrect) pronouncement. The indefinite category is a heterogeneous group and includes those cases where there might well be an underlying malignancy but with insufficient tissue sampling for a more definite diagnosis, cases with atypia adjacent to ulceration which may or may not be malignant, and lower grade atypia where features are not sufficiently clear-cut for a definite dysplasia diagnosis. It is therefore important to qualify this category with the reason for doubt. This, together with the clinical and endoscopic features, will often point the way to the optimal management in these different scenarios.

We also took the opportunity to re-examine the cases which caused the greatest degree of disagreement between pathologists, and although some of these cases are by their nature very difficult, some consistent features emerged. For example, one common cause of disagreement was the presence of pseudostratified pencil-like nuclei, often extending to the surface, which some interpreted as LGD. These cases never seemed to

progress, and we recommend that these changes are accepted as a regenerative phenomenon in the absence of more obvious nuclear atypia. There was also a difference in the weighting given to inflammation and ulceration in downgrading nuclear atypia. Although ulceration can certainly cause regenerative atypia in nearby epithelium, when severe nuclear atypia is present in the presence of ulceration this often reflects a nearby invasive lesion and a diagnosis of HGD should not necessarily be downgraded in these circumstances (Figure 4). On the other hand, lesser degrees of nuclear atypia may easily overlap with regeneration, and LGD should probably not normally be diagnosed in the presence of ulceration or severe inflammation, but the indefinite category may be used with advice to rebiopsy soon. Surface maturation was another criterion that was interpreted differently. Although this is a useful criterion in recognizing regenerative change, we believe that the absence of apparent surface involvement should not deflect from a diagnosis of dysplasia if significant nuclear atypia is present in the crypts (Figure 7). In this context, immunopositivity for p53 confined to the crypts showing the nuclear atypia is good supportive evidence for a dysplasia diagnosis (Figure 7C). In support of this, in many cases where surface involvement was present, there were also areas showing apparent maturation of similarly dysplastic crypts. This is in agreement with previous evidence, which showed similar molecular changes in crypt dysplasia to those found in conventional dysplasia.[19]

The distinction of HGD from LGD remains a difficult area for pathologists. Current management schemas often suggest a big difference in approach to HGD versus LGD, with oesophagectomy for the former and continued surveillance for the latter.[4] In this study 41% of LGD progressed, and with the recent introduction of EMR, local excision is a viable alternative to oesophagectomy. This might suggest that when a definite diagnosis of dysplasia (LGD or HGD) is made by consensus amongst specialist GI pathologists, an attempt should be made to define and locally

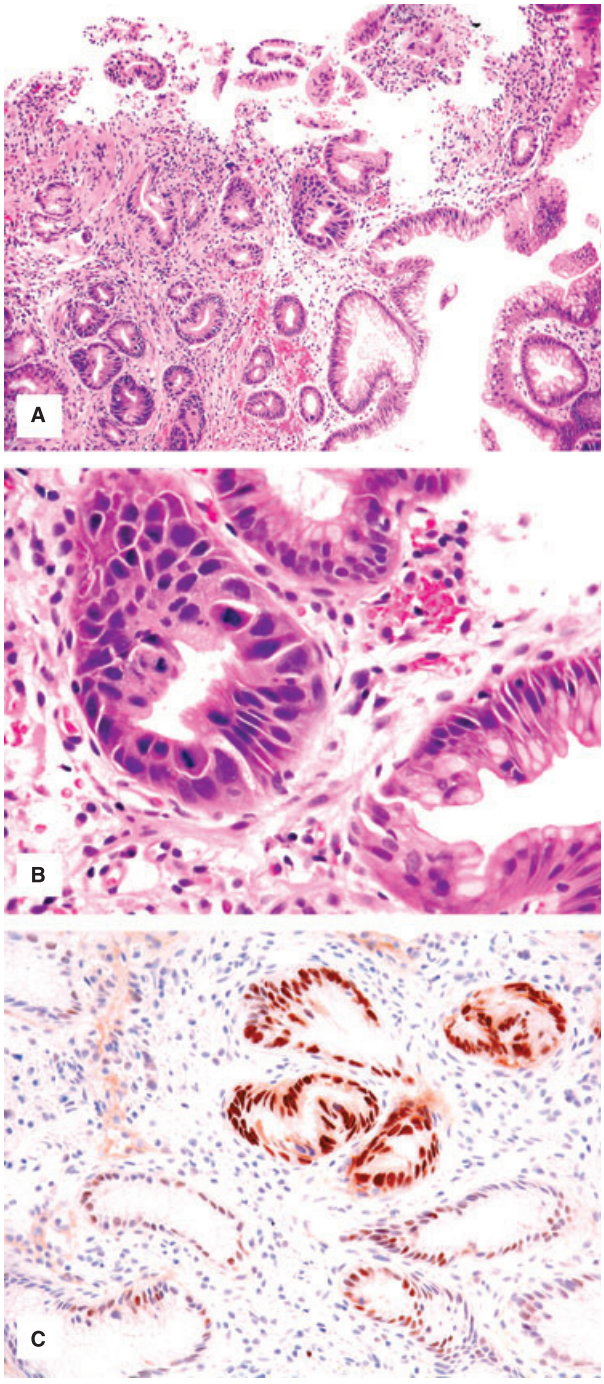

Figure 7. **A**,**B**, Cryptal dysplasia without convincing surface involvement. Note severe nuclear atypia. **C**, p53 on same case showing strong positivity confined to atypical crypts.

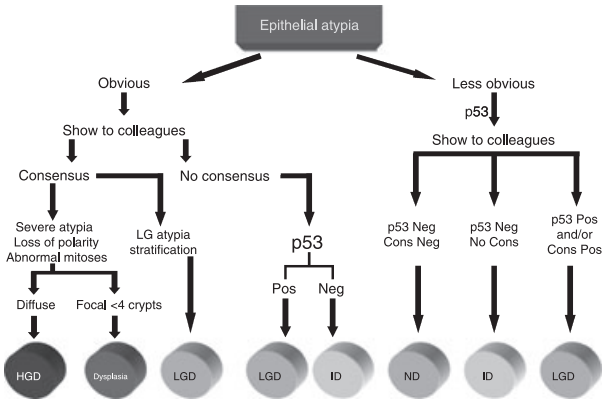

Figure 8. Suggested algorithm for handling atypia in Barrett's oesophagus.

excise the abnormal area as the risk of progression is high. Only where detailed investigation, including endoscopic ultrasound or local excision, shows more advanced disease is oesophagectomy necessarily warranted.

In summary, this study has again reaffirmed the predictive nature of dysplasia in BO when diagnosed using strictly applied criteria by specialist pathologists. It has highlighted the importance of consensus diagnosis for all grades of dysplasia and has shown that LGD when diagnosed in this way is an important diagnosis with a high rate of progression and which should be taken very seriously and assiduously followed up if not treated by local techniques. Furthermore, p53 can play a role in helping refine those indefinite cases that have highest risk of progression. Odze has suggested a helpful algorithm for dysplasia diagnosis in a recent review.[20] Although this is potentially useful, we suggest also considering an additional practical algorithm based on the findings of this study (Figure 8).

Acknowledgements

The authors thank Mr Neil Hand for excellent work in producing p53-stained sections and Miss Anne Kane for assistance with figures and diagrams.

References

1. Haggitt RC. Barrett's esophagus, dysplasia and adenocarcinoma. *Hum. Pathol.* 1994; **25;** 982–993.
2. Skacel M, Petras RE, Gramlich TJ *et al.* The diagnosis of low grade dysplasia in Barrett's oesophagus and its implications for disease progression. *Am. J. Gastroenterol.* 2000; **95;** 3383–3387.
3. Sharma P, Falk GW, Weston AP, Reker D, Johnston M, Sampliner RE. Dysplasia and cancer in a large multicenter cohort of patients with Barrett's esophagus. *Clin. Gastroenterol. Hepatol.* 2006; **4;** 566–572.
4. Brithish Society of Gastroenterology. *Guidelines for the diagnosis and management of Barrett's columnar-lined oesophagus*. BSG guidelines. London: British Society of Gastroenterology, 2005.
5. Schlemper RJ, Kato Y, Stolte M. Review of histological classifications of gastrointestinal epithelial neoplasia: differences in diagnosis of early carcinomas between Japanese and Western pathologists. *J. Gastroenterol.* 2001; **36;** 445–456.
6. Riddell RH, Goldman H, Ransohoff DF *et al.* Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. *Hum. Pathol.* 1983; **14;** 931–968.
7. Reid BJ, Haggitt RC, Rubin CE *et al.* Observer variation in the diagnosis of dysplasia in Barrett's esophagus. *Hum. Pathol.* 1988; **19;** 166–178.
8. Montgomery E, Bronner MP, Goldblum JR *et al.* Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. *Hum. Pathol.* 2001; **32;** 368–378.
9. Kerkhof M, van Dekken H, Steyerberg EW *et al.* CYBAR study group. Grading of dysplasia in Barrett's oesophagus: substantial interobserver variation between general and gastrointestinal pathologists. *Histopathology* 2007; **50;** 920–927.
10. Rabinovitch PS, Longton G, Blount PL, Levine DS, Reid BJ. Predictors of progression in Barrett's esophagus III: baseline flow cytometric variables. *Am. J. Gastroenterol.* 2001; **96;** 3071–3083.
11. Fahmy M, Skacel M, Gramlich TL *et al.* Chromosomal gains and genomic loss of p53 and p16 genes in Barrett's esophagus detected by fluorescence *in situ* hybridization of cytology specimens. *Mod. Pathol.* 2004; **17;** 588–596.
12. Polkowski W, van Lanschot JJ, Ten Kate FJ *et al.* Value of p53 and Ki67 as markers for tumour progression in the Barrett's dysplasia–carcinoma sequence. *Surg. Oncol.* 1995; **4;** 163–171.
13. Dorer R, Odze RD. AMACR immunostaining is useful in detecting dysplastic epithelium in Barrett's esophagus, ulcerative colitis, and Crohn's disease. *Am. J. Surg. Pathol.* 2006; **30;** 871–877.
14. Montgomery E, Goldblum JR, Greenson JK *et al.* Dysplasia as a predictive marker for invasive carcinoma in Barrett esophagus: a follow-up study based on 138 cases from a diagnostic variability study. *Hum. Pathol.* 2001; **32;** 379–388.
15. Skacel M, Petras RE, Rybicki LA *et al.* p53 expression in low grade dysplasia in Barrett's esophagus: correlation with interobserver agreement and disease progression. *Am. J. Gastroenterol.* 2002; **97;** 2508–2513.
16. Weston AP, Banerjee SK, Sharma P, Tran TM, Richards R, Cherian R. p53 protein overexpression in low grade dysplasia (LGD) in Barrett's esophagus: immunohistochemical marker predictive of progression. *Am. J. Gastroenterol.* 2001; **96;** 1355–1362.
17. Murray L, Sedo A, Scott M *et al.* TP53 and progression from Barrett's metaplasia to oesophageal adenocarcinoma in a UK population cohort. *Gut* 2006; **55;** 1390–1397.
18. Lörinc E, Jakobsson B, Landberg G, Veress B. Ki67 and p53 immunohistochemistry reduces interobserver variation in assessment of Barrett's oesophagus. *Histopathology* 2005; **46;** 642–648.
19. Lomo LC, Blount PL, Sanchez CA *et al.* Crypt dysplasia with surface maturation: a clinical, pathologic, and molecular study of a Barrett's esophagus cohort. *Am. J. Surg. Pathol.* 2006; **30;** 423–435.
20. Odze RD. Diagnosis and grading of dysplasia in Barrett's oesophagus. *J. Clin. Pathol.* 2006; **59;** 1029–1038.