

## Assessment of dysplasia in colorectal adenomas: an observer variation and morphometric study

LJR BROWN,\* NC SMEETON,† MF DIXON\*

*From the \*Department of Pathology, University of Leeds, and the †General Practice Research Unit, Institute of Psychiatry, London.*

**SUMMARY** Observer variation in the grading of dysplasia in 100 colorectal adenomas has been analysed by kappa statistics. Intraobserver agreement was only 70% and 67% for the two principal observers, and, as would be expected, interobserver agreement was even lower at 59% and 66%. Although the kappa values were significantly different from chance at the 0.1% level, there were substantial disagreements. When the study was extended to four observers, agreement between observer pairings was considerably worse (as low as 34%), and in four pairings the kappa values did not differ significantly from those expected by chance alone even at the 5% level.

In an endeavour to improve agreement we adopted a percentage estimation grading method; but this failed to achieve any improvement when comparing overall grades. The percentage estimates of the two observers, however, showed a highly significant correlation.

To identify the cytological features given most weight by the principal observers in assessing dysplasia we undertook morphometry on 30 adenomas using an image analysis computer. The nuclear to cytoplasmic ratio, variation in nuclear area, and variation in nuclear height above the basement membrane showed significant differences between mild, moderate, and severely dysplastic epithelia. While evaluation of these parameters therefore appears to be most important in the subjective interpretation of dysplasia, this study has shown that such evaluation is poorly standardised between observers and poorly reproduced within observers.

Our finding of poor agreement in the grading of dysplasia in colorectal adenomas has serious implications for the assessment of dysplasia in inflammatory bowel disease, where the added problem of reactive cellular atypia brings greater complexity to these subjective judgments.

The proper recognition and assessment of dysplasia is essential for the histological diagnosis of premalignant states in the gastrointestinal tract. In many instances the dysplastic lesion bears a close resemblance to the cytological abnormalities found in colorectal adenomas. Thus adenomatous dysplasia is the commonest premalignant change in long-standing ulcerative colitis<sup>1</sup> and constitutes one of the two types of dysplasia found in the stomach.<sup>2</sup>

The natural history of dysplasia in the stomach or large intestine has not been adequately explored, but the finding of severe dysplasia is usually believed to signal an inevitable progression to malignancy and is therefore a mandate for surgical eradication or at least close surveillance. In comparison, the

status of mild dysplasia is more nebulous. Its progression through moderate to severe dysplasia, while highly likely from a theoretical standpoint, has not been well documented. The usual clinical response to a diagnosis of mild dysplasia is to keep the patient under longer interval surveillance or, in some circumstances, to ignore it. Therefore, the histopathologist's ability to distinguish between mild and severe grades of dysplasia is of crucial importance.

Colorectal adenomas provide a plentiful source of epithelium showing dysplastic changes, and we have utilised such material to assess observer variation in grading this type of dysplasia. In addition, we have employed computer aided morphometric methods in an attempt to identify which cytological abnormalities are given most emphasis by histopathologists when making these subjective judgments.

## Material and methods

### GRADING

Haematoxylin and eosin stained sections from 100 colorectal adenomata were obtained from the department of pathology files, randomly renumbered, and individually graded as mild, moderate, or severe dysplasia by two pathologists—MFD, a consultant specialising in gastrointestinal pathology, and LJR, a senior registrar, working independently. Both pathologists had familiarised themselves with the published criteria for grading dysplasia in adenomas.<sup>3-5</sup> The slides were then renumbered and were graded again after an interval of two months. In this way four grades were assigned each adenoma. The "final" grade was the grade most frequently allocated. Where there was an equal division between two grades then the more severe grade was allocated.

Detailed study of adenomas reveals considerable heterogeneity in the degree of dysplasia. For example, some adenomas have small foci of severe dysplasia, while the bulk of the polyp is only mildly dysplastic. If we accept the dictum that polyps must be graded according to the most atypical epithelium represented,<sup>5</sup> such adenomas should be classified as severely dysplastic. None the less, where the severely dysplastic epithelium forms only a small proportion of the whole, problems over its recognition and the variable emphasis attached to its finding could be expected to exaggerate observer disagreement. Under these circumstances it might well be that the same observers would agree on the grade exhibited by the majority of the adenoma.

In an attempt to minimise these problems, a further grading was carried out where the two observers estimated the proportions of each adenoma showing the three grades of dysplasia and expressed these as percentages (run 3). In this way the grade given the highest percentage by each observer could be compared.

In order to compare our "informed" grading with routine practice, two other pathologists of consultant standing separately graded a series of 50 adenomas taken at random from the original collection of 100. They were asked to assess the dysplasia as they would in a routine surgical specimen and to make no reference to published reports concerning grading. Their results were compared with the grades given by the principal observers in run 1. Intra and interobserver agreement was examined by kappa statistics (see below).

In addition, a single observer (MFD) noted whether each adenoma was tubular, villous, or tubulovillous and measured the maximum diameter of the polyp in the section to determine if the mor-

phology or size of the lesion had any influence on the assigned grade.

### MORPHOMETRY

New sections cut at a uniform 2  $\mu$ m from 30 of the adenomas were stained with iron haematoxylin. Ten cases had a final grade of mild dysplasia, 12 moderate, and 8 severe. A representative field was chosen in each and viewed through the 100 $\times$  objective of a Leitz Dialux 20 EM microscope. A Leitz camera lucida drawing tube allowed the self illuminated cursor of the digitising tablet of a Reichert-Jung Kontron IBAS I image analyser to be viewed superimposed on the microscopical field. Within each chosen strip of dysplastic epithelium the nuclei of (usually) 30 adjacent cells were traced around using the cursor on the digitising tablet. Points along the path of the cursor were digitised by the computer and stored electromagnetically. A visual display unit allowed a simultaneous depiction of the tracing. In a similar manner the perimeter of the strip of epithelium containing the traced nuclei and the distance between the apex of each nucleus and the basement membrane were digitised. Each measurement was performed three times to reduce tracing error.

From the stored results a series of parameters and coefficients (Table 1) was calculated, and the results were grouped by assigned grade.

### STATISTICAL METHODS

The principles underlying the use of the kappa statistic and the various coefficients which describe disagreement and bias have been fully explained elsewhere in a study on colorectal cancer.<sup>6</sup> To summarise, kappa is a statistic which measures the agreement between two observers while taking chance agreement into account and is given by the formula:

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

where  $p_o$  and  $p_e$  are the observed and expected proportions of agreement.

Since the ability to distinguish severe dysplasia is of prime interest, the levels of agreement given that the first observer allocates severe (conditional agreement)<sup>6</sup> have been examined.

For the morphometric results, Student's *t* test was applied to discern any significant differences between the grades of dysplasia for each of the calculated variables.

## Results

### GRADING

The intraobserver agreement for runs 1 and 2 are

Table 1 Parameters measured for each adenoma

Mean nuclear area
Standard deviation of nuclear areas
Mean nuclear to cytoplasmic ratio
Mean nuclear height above basement membrane
Standard deviation of nuclear heights
Mean form AR*
Standard deviation of form AR
Mean form PE†
Standard deviation of form PE

\*Form AR defines regular or not regular structures

$$= \frac{\text{Area}}{\pi/4 \times \text{major diameter} \times \text{minor diameter}}$$

†Form PE defines the extent to which the structure differs from a circle

$$= \frac{4\pi \times \text{Area}}{\text{perimeter}^2}$$

Circle = 1; ellipse = <1; irregular structures = <1

shown in Table 2. Both observers show kappa values which are significantly greater than chance agreement ( $p < 0.001$ ), but this is only to be expected in histological practice.<sup>7</sup> Using the benchmarks of Landis and Koch<sup>8</sup> the kappa values are only moderate (MFD) and fair (LJRB), and, of more relevance, the levels of disagreement are substantial (30% for MFD, 33% for LJRB).

Interobserver agreement is analysed in Table 3. As would be expected, the proportions of agreement are less than for intraobserver grading and are only fair on Landis and Koch's scale. Agreement increases (with a corresponding rise in kappa) between runs 1 and 2, and the results for the percentage grading method (run 3), where the majority grade given by each observer was compared, occupy an intermediate position. When the percentage estimates for severe dysplasia alone were compared, however, we found a highly significant correlation ( $r = 0.712$ ,  $p < 0.001$ ) between them. This can be accounted for mainly by agreement on the absence of severe dysplasia, which was found in 41 cases.

Table 3 Interobserver agreement in grading 100 adenomas on two occasions (runs 1 and 2) and by a percentage grading method (run 3)

	$P_o$	kappa	95% limits
Run 1	0.59	0.230	0.049, 0.411
Run 2	0.66	0.369	0.197, 0.541
Run 3	0.64	0.268	0.077, 0.459

The estimates for the remaining 59 cases are shown graphically in the Figure. Despite the high overall correlation the relative pessimism of MFD—that is, a tendency to diagnose severe dysplasia more readily—is evident in these results, and several wide discrepancies in the assessment of severe dysplasia are shown.

Table 4 shows the levels of agreement over the grading of 50 adenomas by four observers. The individual proportions of agreement are low for all observer pairings and kappa is significant only at the 5% level between observers A and B and between C and D. When agreement conditional on the first observer allocating a severe grade is examined (Table 5) the picture is even more disquieting, with gross disparities between all observers.

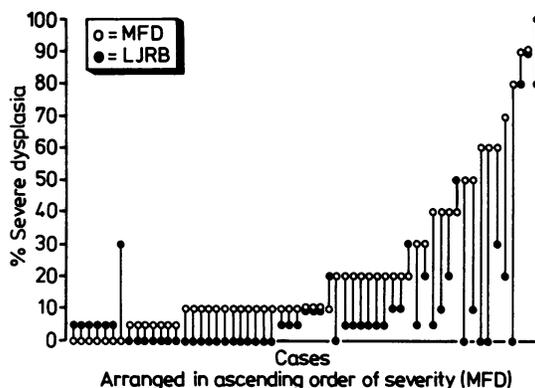
Table 6 shows that when the adenomas are grouped according to their histological type and the mean of the diameters calculated there is an increase in size as the villous component increases. The differences in mean diameter between the tubular, tubulovillous, and villous groups were significant at the 5% level by a one way analysis of variance.<sup>9</sup> Similarly, both the number of allocations of a severe grade (from run 1) and the mean proportion of severely dysplastic epithelium (from run 3) were slightly increased in tubulovillous and villous adenomas (Table 7), but these increases are not statistically significant.

MORPHOMETRY

In addition to the parameters measured directly by

Table 2 Intraobserver agreement in grading 100 adenomas

	MFD Run 1			$p_o$	$k$	95% confidence limits
	Mild	Moderate	Severe			
MFD Run 2						
Mild	5	1	0	0.70	0.450	0.285, 0.615
Moderate	5	44	7			
Severe	1	16	21			
	LJRB Run 1			$p_o$	$k$	95% confidence limits
	Mild	Moderate	Severe			
LJRB Run 2						
Mild	8	6	0	0.67	0.330	0.143, 0.517
Moderate	8	51	11			
Severe	0	8	8			



Graphical representation of the results for the percentage estimates of severe dysplasia by the two principal observers in the 59 adenomas where one or other observer recorded the presence of severe dysplasia.

the computer we have used their standard deviations as a measure of variability and range in each adenoma. All the parameters except nuclear height, form AR (mean and standard deviation), and mean form PE, showed a tendency to increase with increasing grades of dysplasia. The results for the five parameters showing significant differences between grades are shown in Table 8, and their significance values are given in Table 9. Significant differences were seen in the variability of nuclear area, the nuclear to cytoplasmic ratios, and the variability of the heights of the nuclei above the basement membrane (pseudostratification) between all three grades of dysplasia. Other parameters (mean nuclear area, mean nuclear height, and variability in

Table 4 Interobserver agreement between four observers grading 50 adenomas

	<i>P<sub>o</sub></i>	<i>kappa</i>	<i>95% limits</i>
AB	0.54	0.237	0.008, 0.466
AC	0.34	0.038	-0.153, 0.229
AD	0.52	0.164	-0.077, 0.405
BC	0.44	0.144	-0.066, 0.354
BD	0.48	0.115	-0.121, 0.351
CD	0.58	0.278	0.043, 0.513

Table 5 Agreement of second observer given that first allocates as severe (%)

	Second observer			
	A	B	C	D
First observer				
A	—	50	25	13
B	62	—	8	0
C	80	20	—	40
D	100	0	100	—

Table 6 Histological type of adenoma and size

	<i>n</i>	Mean diameter (mm) (SD)
Tubular	53	6.38 (3.8)
Tubulovillous	42	12.66 (6.85)
Villous	5	18.2 (4.4)

Table 7 Relation of severe dysplasias to histological type (MFD)

	<i>n</i>	Severe dysplasia	
		Grade (run 1)	Mean % (run 3)
Tubular	53	13 (24.5%)	11.9
Tubulovillous (42)	47	15 (31.9%)	15.0
Villous (5)			

departure from a circular shape as expressed by the standard deviation of form PE) were significantly different between some but not all grades.

In such a table one would expect one or two significant results at the 5% level on the grounds of chance alone. The number of significant results and the high levels of significance obtained here, however, are very convincing.

### Discussion

The assessment of dysplasia in adenomas of the colorectum is of limited practical value. The complete removal of an adenoma showing even severe dysplasia eliminates the possibility of malignant change in that particular polyp. On the other hand fragments showing only mild or moderate dysplasia may represent the superficial parts of a polypoid adenocarcinoma. Thus in routine diagnostic practice it is much more important to detect invasion beneath the muscularis mucosae and to assess the completeness of removal of the adenomatous epithelium than to recognise the grade of dysplasia within it. Thus in this study we are concerned not so much with the practical implications of dysplasia in adenomas as with the ability of pathologists to recognise degrees of dysplasia and to use this knowledge as a baseline for its recognition in the context of inflammatory bowel disease. Nevertheless, severe dysplasia in colorectal adenoma is a risk factor for the subsequent development of carcinoma and, if found, the patient should undergo regular surveillance.<sup>5</sup> Its detection in adenomas should not therefore be considered purely academic.

We have shown, as have many previous studies,

Table 8 *Morphometric results*

	Mean nuclear area ( $\mu\text{m}^2$ ) ( $\pm$ SD)	SD of nuclear areas ( $\mu\text{m}^2$ ) ( $\pm$ SD)	Mean nuclear to cytoplasmic ratio	SD of nuclear height	SD of Form PE
<i>Dysplasia</i>					
Mild	51.16 $\pm$ 12.21	16.48 $\pm$ 3.18	0.37 $\pm$ 0.26	0.207 $\pm$ 0.096	0.0918 $\pm$ 0.0347
Moderate	52.21 $\pm$ 5.93	20.35 $\pm$ 3.11	0.71 $\pm$ 0.42	0.538 $\pm$ 0.296	0.1164 $\pm$ 0.0173
Severe	74.76 $\pm$ 11.65	34.53 $\pm$ 9.58	1.47 $\pm$ 0.53	1.136 $\pm$ 0.407	0.142 $\pm$ 0.0178

SD = standard deviation.

that adenomas in which there is a villous component are generally larger than the purely tubular type and that there is a tendency for the severity of dysplasia to be linked to both size and villous pattern. Previous studies have not, however, adequately examined observer agreement on the recognition of severe dysplasia.

Our results for the assessment of the overall grade of dysplasia are disquieting in that they show wide disagreements between observers. Even the principal observers, who by virtue of training and preliminary discussions might be expected to employ closely similar standards, agreed on the grade in only 59% (run 1) and 66% (run 2) of cases. When extended to four observers, four of the six observer pairings failed to achieve significant agreement even at the 5% level. The agreement of a second observer given that the first allocated a grade of severe showed even more discrepancies; one observer pairing never agreed on the allocation of a severe grade in 50 adenomas.

A further grading by the principal observers based on an estimation of the proportion of each category of dysplasia present in 100 polyps did not show any improvement when the majority grade was compared ( $p_0 = 0.64$ ,  $k = +0.268$ ), although agreement on the presence or absence of severe dysplasia rose to 69%.

We believe that computerised image analysis has not been used previously on histological sections to measure cytological parameters in colorectal adenomas. In two other studies on colonic adenomas one<sup>10</sup> used stereology to elucidate their microanatomy and the other<sup>11</sup> concentrated mainly

on nuclear and nucleolar size using computerised analysis of electron microscope images. By using standard histological material we hoped to assess which parameters are given most weight in assigning grade in routine diagnostic practice. Although morphometry was performed on thin paraffin sections, we will have overestimated nuclear area by failing to distinguish between overlapping structures, but in a comparative study such as this, correction factors are not necessary.<sup>12</sup> Our values of nuclear area are larger than those of Sato *et al.*,<sup>11</sup> presumably due to the absence of overlapping structures in ultrathin sections.

It is interesting that the mean nuclear height above the basement membrane is not a particularly useful discriminant between grades; this apparently contradicts the advice of Kozuka.<sup>4</sup> It is easy to see, however, how the mean of multiple measurements of different nuclear heights throughout the whole height of the epithelium could give a value not significantly different from a series of nearly identical heights at the midpoint of the epithelium. In this case it is the variability of nuclear height (standard deviation of the nuclear height) which gives significant differences between all three grades.

On the basis of our morphometric results it would seem that the nuclear to cytoplasmic ratio is, as would be expected, important in grading dysplasia and the variability of nuclear area and variability of nuclear height are more important than absolute nuclear area and height. Departure of nuclear shape from a circular profile is of limited use.

In conclusion, subjective assessment by histopathologists of the grade of adenomatous dys-

Table 9 *Student's t test between results grouped by assigned grade*

	Mild: moderate	Moderate: severe	Mild: severe
Mean nuclear areas	NS	<0.0001	<0.01
SD nuclear areas	<0.02	<0.01	<0.01
Mean nuclear to cytoplasmic ratio	<0.05	<0.01	<0.001
Mean nuclear heights	<0.01	NS	NS
SD nuclear heights	<0.005	<0.005	<0.005
SD form PE	NS	<0.01	<0.01

NS = not significant.

SD = standard deviation.

### Assessment of dysplasia in colorectal adenomas

plasia appears to be based on their evaluation of the nuclear to cytoplasmic ratios, variations in nuclear size, and variations in height above the basement membrane seen in an epithelium. This study has shown wide disagreements between observers in making these assessments, especially when assigning an overall grade to an adenoma. It is possible, however, to improve agreement when severe dysplasia is specifically sought. Our findings underline the difficulties faced by histopathologists in the assessment of premalignant cytological changes and explain why attempts to standardise the classification of dysplasia in the more difficult context of inflammatory bowel disease<sup>13</sup> have met with only limited success in eliminating disagreements even between expert observers.

We are most grateful to Dr Jan Lowe and Dr Christine Bates for participating in this study. We thank Andrew Hay for preparing the figure and Helen Swiercz for typing the manuscript.

#### References

- <sup>1</sup> Riddell RH. The precarcinomatous phase of ulcerative colitis. In: Morson BC, ed. *Current topics in pathology*. New York: Springer-Verlag, 1976;179–219.
- <sup>2</sup> Jass JR. A classification of gastric dysplasia. *Histopathology* 1983;7:181–93.
- <sup>3</sup> Ekelund G, Lindstrom C. Histopathological analysis of benign polyps in patients with carcinoma of the colon and rectum. *Gut* 1974;15:654–63.
- <sup>4</sup> Kozuka S. Premalignancy of the mucosal polyp in the large intestine: I. Histological gradation of the polyp on the basis of epithelial pseudostratification and glandular branching. *Dis Colon Rectum* 1975;18:483–93.
- <sup>5</sup> Konishi F, Morson BC. Pathology of colorectal adenomas: a colonoscopic survey. *J Clin Pathol* 1982;35:830–41.
- <sup>6</sup> Thomas GDH, Dixon MF, Smeeton NC, Williams NS. Observer variation in the histological grading of rectal carcinoma. *J Clin Pathol* 1983;36:385–91.
- <sup>7</sup> Silcocks PBS. Measuring repeatability and validity of histological diagnosis—a brief review with some practical examples. *J Clin Pathol* 1983;36:1269–75.
- <sup>8</sup> Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- <sup>9</sup> Davies OD. *Statistical methods in research and production*. London: Oliver and Boyd, 1947.
- <sup>10</sup> Elias H, Hyde DM, Mullens RS, Frederick C, Lambert BS. Colonic adenomas: stereology and growth mechanisms. *Dis Colon Rectum* 1981;24:331–42.
- <sup>11</sup> Sato E, Goto M, Nakamura T. Nuclear ultrastructure in carcinoma, adenoma, Peutz-Jegher's polyp and dysplasia of the large bowel: A morphometric analysis. *Gann* 1981;72:245–53.
- <sup>12</sup> Franklin CD, Smith CJ. Stereological analysis of histological parameters in experimental premalignant hamster cheek epithelium. *J Pathol* 1980;130:201–15.
- <sup>13</sup> Riddell RH, Goldman H, Ransohoff DF, *et al.* Dysplasia in inflammatory bowel disease: standardized classification with provisional clinical applications. *Hum Pathol* 1983;14:931–68.

Requests for reprints to: Dr MF Dixon, Department of Pathology, University of Leeds, Leeds LS2 9JT, England.