

## ORIGINAL ARTICLE

## How could pathologists improve the initial diagnosis of colitis? Evidence from an international workshop

E Bentley, D Jenkins, F Campbell, B Warren

*J Clin Pathol* 2002;**55**:955–960

**Background:** The taking of multiple colorectal biopsies is in widespread use although there is little research into their benefit for the pathological diagnosis of inflammatory bowel disease. There is also still debate about appropriate morphological criteria for interpreting these biopsies.

**Aims:** To determine the effect of single versus multiple biopsies on the accuracy of diagnosis and to study the accuracy and reproducibility of the different criteria used in the diagnosis of multiple biopsies by expert and non-expert pathologists.

**Method:** Thirteen expert and 12 non-expert international diagnostic histopathologists attended a workshop. Sixty cases with full follow up were viewed, blinded, in two rounds. Diagnoses were made on rectal biopsies and then full colonoscopic series.

**Results:** Experts correctly identified 24% of Crohn's disease cases (non-experts, 12%) from the rectal biopsies. This improved to 64% (non-experts, 60%) with the full series. The accuracy of the diagnosis of ulcerative colitis also improved slightly with the full series from 64% to 74% overall. Experts had a similar (moderate) level of agreement and accuracy to non-experts. For Crohn's disease, the likelihood ratios (LR) for the most important individual features were 12.4 for granulomas and 3.3 for focal or patchy inflammation. Features favouring ulcerative colitis were diffuse crypt architectural irregularity (LR, 3.4), general crypt epithelial polymorphs (LR, 3.7), and reduced crypt numbers (LR, 2.9).

**Conclusions:** A full colonoscopic series gave more accurate diagnosis than a rectal biopsy. Accurate pathologists used the same evidence based criteria for multiple biopsies as for single biopsies.

See end of article for authors' affiliations

Correspondence to: Dr D Jenkins, Division of Pathology, Queen's Medical Centre, Nottingham NG7 2UH, UK; david.jenkins@nottingham.ac.uk

Accepted for publication 16 July 2002

**B**iopsy diagnosis is a crucial step for the clinical management of suspected inflammatory bowel disease. However, in many centres worldwide there is wide variation in clinical practice for colorectal biopsy, with some using multiple biopsies<sup>1</sup> and others using a single rectal biopsy.<sup>2</sup>

Many different criteria are used to interpret these biopsies and may contribute to diagnostic variation.<sup>3</sup> In 1997, the British Society of Gastroenterology published guidelines for the initial biopsy diagnosis of suspected chronic inflammatory bowel disease.<sup>4</sup> These represent a systematic review using evidence based methods into the use of a single colorectal biopsy. Only a small proportion of suggested criteria were found to be reproducible and accurate. Eligible criteria were found to have moderate reproducibility, with a  $\kappa$  value of at least 0.4, or percentage agreement of at least 80%. They also had to achieve diagnostic sensitivity and specificity of 50% or more in at least one study.

In this review,<sup>4</sup> criteria favouring a diagnosis of ulcerative colitis (UC) on a single colorectal biopsy were crypt architectural distortion and diffuse transmucosal inflammation. Those favouring a diagnosis of Crohn's disease (CD) were discontinuous crypt distortion and discontinuous inflammation. Granulomas were found to be very specific, but not very sensitive, features of CD; a finding confirmed by Tobin *et al.*<sup>5</sup>

"In many centres worldwide there is wide variation in clinical practice for colorectal biopsy, with some using multiple biopsies and others using a single rectal biopsy"

Multiple biopsies are now in widespread use; one recent study has suggested that diagnostic accuracy is higher,<sup>1</sup> but there are few formal studies comparing the accuracy, reproducibility and criteria with those for single biopsies. Another aspect that could provide valuable information is the pattern of features between biopsies. For example, finding a granuloma in one of

several biopsies would favour a diagnosis of CD, whereas more widespread distribution of diffuse inflammation might favour UC.

In major specialist centres, biopsies are often interpreted by expert gastrointestinal pathologists. In community or district general hospitals, diagnosis is usually part of general histopathology practice. It is possible that experts may use criteria that have not yet been formally identified.

The aims of this workshop were to study the contributions of multiple and single biopsies, expert status, brief exposure to guidelines, and the use of particular evidence based diagnostic criteria to the accuracy and the reproducibility of diagnosis of intestinal inflammation. The implications of these results for training in diagnostic pathology are discussed.

## MATERIALS AND METHODS

Diagnostic histopathologists from Europe, North America, and from other parts of the world were invited to attend a workshop held in Nottingham, UK during July 2000. The final group consisted of 13 experts and 12 non-experts from France, Belgium, Germany, Holland, Canada, Sri Lanka, Japan, USA, and the UK.

An expert gastrointestinal pathologist was defined as someone having: membership of a professional organisation devoted to gastrointestinal pathology; published work on the pathology of inflammatory bowel disease; a diagnostic practice of at least 1000 gastrointestinal specimens each year; and at least five years of specialist gastrointestinal pathology. A general pathologist or non-expert was defined as one: practising in a community/district general hospital practice; who

**Abbreviations:** CD, Crohn's disease; LR, likelihood ratio; UC, ulcerative colitis

**Table 1** Number\* and percentage (in parenthesis) of patients with Crohn's disease and ulcerative colitis (UC) correctly diagnosed by 13 experts and 12 non-experts on the rectal biopsies and full series in rounds 1 and 2

|             | Crohn's disease |               | UC            |               |
|-------------|-----------------|---------------|---------------|---------------|
|             | Experts         | Non-experts   | Experts       | Non-experts   |
| Round 1     |                 |               |               |               |
| Rectal      | 39 (17) [18]    | 31 (14) [10]  | 190 (65) [21] | 170 (62) [15] |
| Overall     |                 | 70 (16)       |               | 360 (64)      |
| Full series | 130 (56) [15]   | 109 (50) [10] | 213 (74) [24] | 195 (72) [16] |
| Overall     |                 | 239 (53)      |               | 408 (73)      |
| Round 2     |                 |               |               |               |
| Rectal      | 58 (24) [8]     | 26 (12) [7]   | 195 (66) [15] | 170 (62) [14] |
| Overall     |                 | 84 (18)       |               | 365 (64)      |
| Full series | 152 (64) [9]    | 131 (60) [10] | 212 (73) [22] | 203 (76) [20] |
| Overall     |                 | 283 (62)      |               | 415 (74)      |

\*Although there were 19 cases of CD and 24 of UC, not all cases were viewed by all pathologists; missing values are shown in square brackets. The results are presented as a percentage of the total diagnoses available for each category. If all cases had been viewed, for CD n=247 (experts) and 228 (non-experts). If all UC cases had been viewed n=312 (experts) and 288 (non-experts).

had no nationally declared interest/involvement in gastrointestinal pathology or publications, and whose involvement with gastrointestinal pathology was less than 40% of the workload.

At the start of the meeting, participants were asked to complete a questionnaire about their background and practices, and a questionnaire about what they considered to be the important diagnostic histological features of UC, CD, infective colitis, and microscopic colitis.

The format of the workshop was a first round of case viewing, followed by a discussion about features and evidence based guidelines. For the second round the cases were renumbered in random order and viewed again. All cases were viewed blinded to the clinical history. The meeting ended with a discussion about the cases and an open general review of cases in relation to their outcome.

There were 60 cases for viewing. All these cases were initial colonoscopic series and had at least five years of follow up of clinical and pathological outcomes, so that a definitive diagnosis could be given. Cases consisted of a rectal or rectosigmoid biopsy slide plus four to six (median, five) slides from other sites in the colon and ileum. There were eight possible categories of diagnosis with 19 cases of CD, 24 of UC, five normal, four each of indeterminate and infective colitis, and four other cases consisting of two collagenous colitis and one case each of tuberculosis and graft versus host disease.

Data collection sheets were completed for each case. There was space to indicate which features were important for the particular slide, and whether a feature helped make or change the diagnosis. Pathologists were asked to give a diagnosis after viewing the first rectal slide, then a further diagnosis after viewing the full colonoscopic series. Participants were identified by randomly allocated number only.

The choices of diagnosis were CD, UC, normal, or other, with a free text field for the "other" diagnosis. Where both CD and UC had been selected, but no other diagnosis, this was coded as indeterminate. With any other combinations the stated "other" diagnosis—for example, collagenous or infective colitis—was used if available. If not, the diagnosis was coded as "other". Individual features were recorded as present or absent. Twenty eight features were used for the first round. During the discussion session after round one, eight were excluded from the list or combined, and three relating to the terminal ileum were added by request of the participants. This gave a choice of 23 features for the second round. Because of time constraints, not all cases were viewed by all pathologists

in each round. The 32 cases viewed by all 25 pathologists (complete cases) in each round were used for comparing the features used in the second round. With regard to the features, we compared cases that were correctly diagnosed, regardless of the type of pathologist.

Pathologists were asked to choose which features were important for the diagnosis. For each feature the number of complete cases, with 50% or more of all pathologists selecting that feature, was determined. Likelihood ratios (LR) were calculated for each individual feature, to determine whether they favoured CD or UC. These express the odds that a feature is present in a case with (as opposed to without) the target disorder. Good discriminatory odds are less than 0.1 and greater than 10.0.<sup>6</sup>

Interobserver agreement was assessed using the  $\kappa$  statistic. Standard errors were estimated by bootstrap resampling, using 200 samples and the Resampling Stats EXCEL add-in. This allows for non-independence in the observations (for example, in comparing the effect on agreement of having the full series as well). A  $\kappa$  score of 0.4 is considered moderate agreement.<sup>7</sup> Other statistics used were  $\chi^2$  for categorical comparisons and Mann-Whitney U for comparisons between non-parametric ordinal data.

## RESULTS

The second round represented the optimal achievement of all pathologists and the results for this are presented in detail. Differences between first and second rounds are presented where appropriate.

### Multiple biopsies improve accuracy of diagnosis

Table 1 compares the percentage of diagnoses, from cases with a final clinical pathological reference diagnosis of CD or UC, correctly identified by experts and non-experts using the rectal biopsies and full series. Results were similar for both rounds. In the second round only 18% of reference CD cases were correctly diagnosed overall on the rectal biopsies, whereas 62% were correct with the full series. The consensus diagnosis for CD cases given by most of the 25 participants using the full series was CD in 15 of 19 cases, UC in two cases, and one each of infective colitis and normal. Total agreement by all pathologists on a case using the full series was achieved for only one case by experts and for two cases by non-experts.

In contrast, the percentage of UC cases correctly diagnosed improved only from 64%, using the rectal biopsies, to 74% with the full series in the second round, almost identical to the first

**Table 2** Features used: number of cases where 50% or more of all pathologists correctly identifying each case selected feature as important and likelihood ratios (LR) for features distinguishing between Crohn's disease (CD) and ulcerative colitis (UC)

|  | CD (n=12)       |      | UC (n=12)       |     |
|--|-----------------|------|-----------------|-----|
|  | Number of cases | LR   | Number of cases | LR  |
| <b>Architecture</b>                          |                 |      |                 |     |
| (1) Focal crypt architectural irregularity   | 8               | 1.7  | 2               | 0.6 |
| (2) Diffuse crypt architectural irregularity | 1               | 0.3  | 11              | 3.4 |
| (3) Reduced crypt numbers/atrophy            | 2               | 0.3  | 9               | 2.9 |
| <b>Chronic inflammation</b>                  |                 |      |                 |     |
| (4) Focal or patchy                          | 8               | 3.3  | 1               | 0.3 |
| (5) Basal plasma cells                       | 5               | 0.4  | 12              | 2.3 |
| (6) Diffuse superficial inflammation         | 1               | 0.6  | 0               | 1.1 |
| (7) Diffuse transmucosal inflammation        | 7               | 0.5  | 12              | 1.8 |
| (8) Granulomas (mucin)                       | 0               | 2.0  | 0               | 0.5 |
| (9) Granulomas (epithelioid)                 | 7               | 12.4 | 0               | 0.1 |
| <b>Polymorph inflammation</b>                |                 |      |                 |     |
| (10) Lamina propria polymorphs               | 10              | 0.9  | 11              | 1.2 |
| (11) Focal crypt epithelial polymorphs       | 9               | 1.1  | 6               | 1.0 |
| (12) General crypt epithelial polymorphs     | 1               | 0.3  | 3               | 3.7 |
| (13) Crypt abscess                           | 7               | 0.6  | 10              | 1.8 |
| (14) Polymorph exudate                       | 1               | 0.6  | 4               | 1.6 |
| <b>Epithelial changes</b>                    |                 |      |                 |     |
| (15) Erosion/ulceration                      | 4               | 1.2  | 3               | 0.7 |
| (16) Mucin depletion                         | 3               | 0.6  | 9               | 1.7 |
| (17) Paneth cells distal to hepatic flexure  | 0               | 0.7  | 0               | 1.8 |
| <b>Epithelial associated changes</b>         |                 |      |                 |     |
| (18) Increased intraepithelial lymphocytes   | 0               | 2.5  | 0               | 0.4 |
| (19) Increased subepithelial collagen        | 0               | 1.0  | 0               | 1.0 |

round. For UC, the consensus diagnosis agreed with the reference standard in 21 of 24 cases, with the three remaining cases being diagnosed as CD. Total agreement by all the pathologists was achieved for three cases by the experts, whereas all the non-experts agreed on two cases.

With the normal cases, rectal biopsies and the full series gave similar results. On average, 61% of experts and 62% of non-experts correctly diagnosed these cases on the rectal biopsies, whereas for the full series the figures were 57% and 66%, respectively. The consensus diagnosis for all five cases was normal.

#### Evidence based criteria apply to multiple biopsies

There were 12 cases of Crohn's disease and 12 cases of UC for which the full series were viewed by all 25 pathologists (table 2). In the diagnosis of CD, the most discriminant feature was epithelioid granulomas, with an LR of 12.4. This was selected in seven of 12 cases by over half the pathologists who correctly identified each case. Other important features were focal or patchy inflammation (LR, 3.3), selected in eight of 12 cases. Some features were selected in a high number of cases, but had low LRs. For example, focal crypt architectural irregularity was selected in eight cases but had an LR of 1.7, whereas lamina propria polymorphs were selected in 10 cases but had an LR of 0.9.

The most discriminant features for UC were diffuse crypt architectural irregularity (LR, 3.4), selected in 11 of 12 cases, reduced crypt numbers (LR, 2.9) in nine of 12 cases, and basal plasma cells (LR, 2.3) in all cases. Diffuse crypt epithelial polymorphs had an LR of 3.7, but was only selected in three of 12 cases. Diffuse transmucosal inflammation, lamina propria polymorphs, and crypt abscesses were all selected in most cases, but had LRs of less than two.

#### Multiple biopsies increase yield of focal CD lesions

In 10 of 19 cases of CD the consensus diagnosis from the rectal biopsy was normal. The average number of pathologists changing diagnosis for each case after viewing the full series was 11 for CD and three for UC. Experts and non-experts were

equally likely to change. The most important feature selected for changing to CD after viewing multiple biopsies was submucosal epithelioid granulomas, selected by both experts and non-experts, in a quarter of the cases. Other features considered important for the diagnosis of CD on multiple biopsies were focal crypt architectural irregularity, diffuse crypt architectural irregularity, and patchy inflammation.

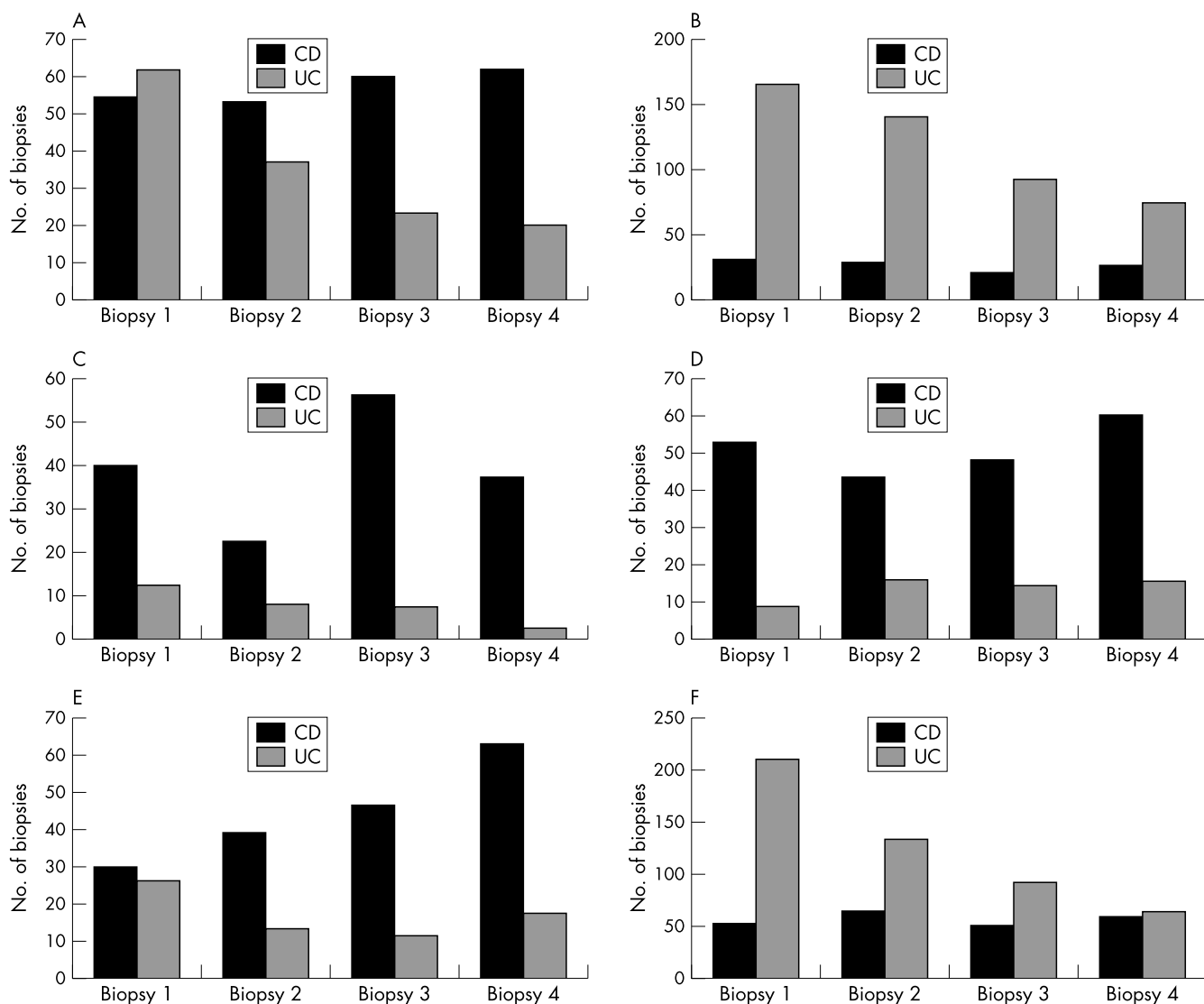
There was no simple uniform pattern of interpretation of multiple biopsies, with a wide variation in features selected by individual pathologists. However, there were some trends that could be discerned (fig 1). For UC, most pathologists reported a pattern of more frequent abnormalities distally and a trend towards a decrease proximally for both focal and diffuse crypt architectural irregularity and for diffuse transmucosal inflammation. For CD, there was no clear pattern of involvement between proximal and distal biopsies, except for patchy inflammation, noted more frequently in proximal biopsies.

#### Expertise and exposure to guidelines

Experts and non-experts gave similar accuracy when diagnosing both CD and UC. There was very little difference between round 1 and 2 for UC (table 1). The percentages of UC cases correctly diagnosed by experts were 74% and 73%, and for non-experts 72% and 76%, for rounds 1 and 2, respectively. Both experts and non-experts improved in diagnosing CD after discussing the guidelines. The percentages of CD cases correctly diagnosed by experts were 56% and 64%, and for non-experts 50% and 60%, for rounds 1 and 2, respectively.

#### Reproducibility

Reproducibility was poor for CD on the rectal biopsies ( $\kappa = 0.18$  for both experts and non-experts) and only achieved moderate agreement with the full series ( $\kappa = 0.43$  and  $0.38$ , respectively; table 3). Both experts and non-experts were more likely to agree about the diagnosis of UC. On the rectal biopsies,  $\kappa$  scores were 0.56 and 0.39, with slightly better agreement on the full series of 0.64 for experts and 0.53 for non-experts, respectively.



**Figure 1** Number of biopsies with selected feature in cases with a consensus diagnosis of Crohn's disease (CD) and ulcerative colitis (UC). The selected features are: (A) focal crypt architectural irregularity; (B) Diffuse crypt architecture irregularity; (C) granulomas; (D) focal inflammation; (E) patchy inflammation; and (F) diffuse transmucosal inflammation. Biopsy 1 is the rectal biopsy, and biopsies 2 to 4 are progressively more proximal.

Interobserver agreement was moderate for the normal cases ( $\kappa = 0.47$  experts, 0.49 non-experts) and for collagenous colitis ( $\kappa = 0.57$  experts, 0.43 non-experts; table 3). There was virtually no interobserver agreement for indeterminate colitis. The consensus diagnosis for these four cases was CD for three cases and collagenous colitis for one case. There was no useful level of interobserver agreement on the categories of infective colitis, graft versus host disease, and tuberculosis.

**DISCUSSION**

Our study demonstrates the value of multiple colonoscopic biopsies and the importance of evidence based criteria in the initial diagnosis of colitis. It also showed that experts did not offer more accurate diagnoses than non-experts or possess special criteria for better diagnosis.

There is limited research on the use of multiple biopsies, despite this being common practice. The best diagnostic

**Table 3** Kappa scores (95% confidence interval) for experts and non-experts, for each diagnostic category and overall, using the second round rectal biopsy alone and the full colonoscopic series

|                    | Experts (rectal)     | Non-experts (rectal) | Experts (full series) | Non-experts (full series) |
|--------------------|----------------------|----------------------|-----------------------|---------------------------|
| Crohn's disease    | 0.18 (-0.01 to 0.38) | 0.18 (-0.01 to 0.38) | 0.43 (0.32 to 0.53)   | 0.38 (0.28 to 0.49)       |
| Ulcerative colitis | 0.56 (0.45 to 0.67)  | 0.39 (0.30 to 0.48)  | 0.64 (0.55 to 0.74)   | 0.53 (0.44 to 0.63)       |
| Indeterminate      | 0.09 (0.03 to 0.15)  | 0.06 (0.01 to 0.11)  | 0.07 (0.02 to 0.11)   | -0.03 (-0.12 to 0.05)     |
| Normal             | 0.57 (0.45 to 0.69)  | 0.55 (0.45 to 0.64)  | 0.47 (0.24 to 0.71)   | 0.49 (0.26 to 0.73)       |
| Infective          | 0.21 (-0.02 to 0.43) | 0.11 (0.04 to 0.18)  | 0.28 (0.06 to 0.50)   | 0.14 (0.03 to 0.25)       |
| Collagenous        | 0.52 (0.14 to 0.90)  | 0.41 (0.10 to 0.73)  | 0.57 (0.24 to 0.90)   | 0.43 (0.20 to 0.66)       |
| Graft versus host  | 0.05 (0.02 to 0.08)  | 0.02 (-0.02 to 0.07) | 0.19 (0.03 to 0.34)   | 0.15 (0.04 to 0.25)       |
| Tuberculosis       | 0.04 (0.02 to 0.06)  | *                    | -0.02 (-0.08 to 0.05) | *                         |
| Overall            | 0.35 (0.29 to 0.42)  | 0.29 (0.24 to 0.34)  | 0.43 (0.36 to 0.51)   | 0.39 (0.32 to 0.45)       |

\*No diagnosis made in this category.



accuracy, for both experts and non-experts in the study, came from the examination of a full series, particularly for CD. Using a full colonoscopic biopsy series, rather than a single rectal biopsy, produced the largest diagnostic improvement. The percentage of correct diagnoses of CD against a reference standard increased from 12% to 60% for non-experts and 24% to 64% for experts. The diagnostic improvement using multiple biopsies for UC was 7% for experts and 14% for non-experts. Although the clinical relevance of examining rectal biopsies alone in suspected cases of CD may be questioned by some, the provision of a single biopsy is still common practice in some areas. Hence, the importance of establishing the need for a full biopsy series.

“Using a full colonoscopic biopsy series, rather than a single rectal biopsy, produced the largest diagnostic improvement”

The most common diagnosis given for the rectal biopsies, in the cases of CD where the diagnosis was made on the series and not on the rectal biopsy, was “normal”. The presence of focal inflammation, focal architectural irregularity, or epithelioid granulomas in other biopsies helped to determine the diagnosis. Multiple biopsies allowed more accurate diagnoses by confirming the presence of these features. This is consistent with the biological nature of the diseases where CD is patchy and often spares the rectum. Diffuse chronic inflammation within individual biopsies and in multiple sequential biopsies was found to help distinguish UC from CD by Konuma *et al.*<sup>8</sup>

The use of criteria by pathologists who made accurate diagnoses of the two diseases was determined. These criteria were consistent with those identified from previous studies of reproducibility and accuracy, summarised in the British Society of Gastroenterology guidelines.<sup>4</sup> From the data in our current study likelihood ratios for these were calculated. Important features for the diagnosis of CD were granulomas (LR, 12.4) and focal or patchy inflammation (LR, 3.3). Similarly for UC, important features were diffuse crypt epithelial polymorphs (LR, 3.7), diffuse crypt architectural irregularity (LR, 3.4), reduced crypt numbers/atrophy (LR, 2.9), and basal plasma cells (LR, 2.3).

Expertise in this area did not guarantee better reproducibility, or accuracy, because experts were only marginally better than non-experts in our study. It has to be stressed, however, that the aim of this interobserver study was to investigate morphological criteria in controlled viewing conditions, rather than full diagnostic competence per se. The lack of clinical and radiological data supporting the diagnoses may have affected experts more than non-experts, because they may use this additional information with greater sophistication than general pathologists.

One of the areas in which expertise was thought to lie was through the use of unpublished unidentified criteria or interpretation. However, we identified no new criteria, previously confined to expert practice. It is possible that non-experts improved their performance because they were provided with a template and discussion of criteria that helped them, whereas the experts were already more familiar with the criteria. The role of the expert should be to provide the best evidence based descriptions of disease so that others can use this information. Discussion of the guidelines in between rounds of viewing marginally improved performance. However, a more formal assessment of the effect of guidelines was not undertaken.

The workshop confirmed previous findings that both the reproducibility and accuracy of diagnosis were only moderately good for experts and non-experts. Reproducibility was moderate or poor ( $\kappa < 0.6$ ) for all diagnostic categories for both experts and non-experts. However, the use of multiple biopsies improved reproducibility compared with rectal

### Take home messages

- A full colonoscopic series gave more accurate diagnosis than a single rectal biopsy, particularly in Crohn's disease (CD), where it improved from 24% correctly diagnosed (non-experts, 12%) to 64% (non-experts, 60%)
- The improvement for the diagnosis of ulcerative colitis (UC) was less pronounced, with an overall improvement from 64% to 74%
- Features favouring CD were granulomas and focal or patchy inflammation, those favouring UC were diffuse crypt architectural irregularity, general crypt epithelial polymorphs, and reduced crypt numbers
- Accurate pathologists used the same evidence based criteria for multiple biopsies as for single biopsies, and when using these criteria the performance of the experts and non-experts was very similar
- The workshop approach is an excellent method to explore, in more detail, the basis of pathological disagreement and develop improved definitions of criteria for intestinal pathology and other areas of diagnostic cellular pathology

biopsies, for UC to 0.64 and 0.53, and for CD to 0.44 and 0.38 for experts and non-experts, respectively. Theodossi and colleagues<sup>9</sup> also found low  $\kappa$  values in their study of single rectal biopsies using seven experts and three non-experts, with values of 0.37 for UC and 0.20 for CD. Poor diagnostic performance was associated with a failure to identify many of the criteria recognised by the good performers in a high proportion of biopsies, but also occurred when a correct identification of the appropriate diagnostic criteria had been made. One of the ways in which reproducibility and the consequent accuracy of pathological diagnosis could be improved would be to use an expert system. This could combine multiple images showing examples of well defined evidence based features to support slide reading, and a logical expert system to support interpretation. This logical expert system could incorporate decision rules, such as those developed in recent studies.<sup>3,10</sup> The introduction of such computerised systems could be more effective than paper guidelines and might help prevent overdiagnosis.<sup>11</sup> A computerised system could be used alongside the microscope during routine pathological reporting, and be used to support the production of the report, to ensure that good practice is automatically followed. Such a system has recently been described for training in breast fine needle aspiration cytology.<sup>12</sup>

An important feature of any diagnostic study is the selection of the reference standard against which the study diagnoses are being compared. For our study, the clinicopathological diagnosis based on five years of follow up was used. This did introduce the possibility that some of the initial biopsies may not have provided sufficient information to make a definitive diagnosis. A measure of this effect is obtained from the observation that a majority consensus diagnosis was not reached in 21% of cases of CD and 12% of UC cases on the full series.

“A computerised system could be used alongside the microscope during routine pathological reporting, and be used to support the production of the report, to ensure that good practice is automatically followed”

The term “indeterminate” colitis has been applied to cases where features do not favour CD or UC. It has been proposed that the term should be used as a “pending tray” diagnosis, representing diagnostic inadequacy, and not as a specific nosological entity.<sup>13</sup> The poor accuracy and reproducibility of this category demonstrated in our study supports this view. The use of a generic term such as non-specific inflammation was allowed through the use of the “other” category. This was

only used in a very small proportion of cases, although when used it was favoured more by non-experts, and then only a minority. Use of the term reduced in the second round after discussion of criteria.

In conclusion, the use of colonoscopic biopsies is essential for the accurate diagnosis of initial inflammatory bowel disease. Accurate pathologists used the same evidence based criteria for multiple biopsies as for single biopsies. Using these criteria the performance of the non-expert pathologist was very similar to the experts. The role of the expert should be to provide the best evidence based descriptions of disease and to show how this can be integrated with radiological and clinical information. To do this requires a rigorous approach using the best methods of diagnostic research.<sup>14</sup> As part of this, the workshop approach would be an excellent method to explore, in more detail, the basis of pathological disagreement and develop improved definitions of criteria for intestinal pathology and other areas of diagnostic cellular pathology.

### ACKNOWLEDGEMENTS

Members of the International workshop: A Borczuk (New York, USA), D Chatelain (Amiens, France), A Clark (Wirral, UK), C Cuvelier (Gent, Belgium), A Driessen (Maastricht, Netherlands), B Fabiani (Le Mans, France), JF Flejou (Paris, France), T Fukuda (Fukushima, Japan), K Geboes (Leuven, Belgium), N Goldstein (West Bloomfield, USA), J Greenson (Michigan, USA), J Hewavisenthi (Ragama, Sri Lanka), A Jouret (Peruwelz, Belgium), S Kurian (Vellore, India), A Lazenby (Alabama, USA), K Lewin (Los Angeles, USA), M Mathan (Dhaka, Bangladesh), N Ostrzega (Beverly Hills, USA), S Perera (Ragama, Sri Lanka), R Riddell (Toronto, Canada), H Rotterdam (New York, USA), M Tanaka (Hirosaki, Japan), A von Herbay (Heidelberg, Germany), H Watanabe (Niigata, Japan), S Yonezawa (Kagoshima, Japan).

This workshop was sponsored and funded by: The International Organisation for the study of Inflammatory Bowel Disease (IOIBD), the Pathological Society of Great Britain and Ireland, and Astra Zeneca.

.....

### Authors' affiliations

**E Bentley, D Jenkins**, Division of Pathology, Queen's Medical Centre, Nottingham, NG7 2UH, UK

**F Campbell**, Department of Pathology, Royal Liverpool University Hospital, Liverpool, L7 8XP, UK

**B Warren**, Department of Cellular Pathology, John Radcliffe Hospital, Oxford, OX3 9DU, UK

### REFERENCES

- 1 **Geboes K**, Ectors N, D'Haens G, *et al*. Is ileoscopy with biopsy worthwhile in patients presenting with symptoms of inflammatory bowel disease? *Am J Gastroenterol* 1998;**93**:201–6.
- 2 **Dundas SA**, Dutton J, Skipworth P. Reliability of rectal biopsy in distinguishing between chronic inflammatory bowel disease and acute self-limiting colitis. *Histopathology* 1997;**31**:60–6.
- 3 **Tanaka M**, Riddell RH, Saito H, *et al*. Morphologic criteria applicable to biopsy specimens for effective distinction of inflammatory bowel disease from other forms of colitis and of Crohn's disease from ulcerative colitis. *Scand J Gastroenterol* 1999;**34**:55–67.
- 4 **Jenkins D**, Balsitis M, Gallivan S, *et al*. Guidelines for the initial biopsy diagnosis of suspected chronic idiopathic inflammatory bowel disease. The British Society of Gastroenterology initiative. *J Clin Pathol* 1997;**50**:93–105.
- 5 **Tobin JM**, Sinha B, Ramani P, *et al*. Upper gastrointestinal mucosal disease in pediatric Crohn disease and ulcerative colitis: a blinded, controlled study. *J Pediatr Gastroenterol Nutr* 2001;**32**:443–8.
- 6 **Simel DL**, Samsa GP, Matchar DB. Likelihood ratios with confidence: sample size estimation for diagnostic test studies. *J Clin Epidemiol* 1991;**44**:763–70.
- 7 **Silcocks PB**. Measuring repeatability and validity of histological diagnosis—a brief review with some practical examples. *J Clin Pathol* 1983;**36**:1269–75.
- 8 **Konuma Y**, Tanaka M, Saito H, *et al*. A study of the histological criteria for ulcerative colitis: retrospective evaluation of multiple colonic biopsies. *J Gastroenterol* 1995;**30**:189–94.
- 9 **Theodossi A**, Spiegelhalter DJ, Jass J, *et al*. Observer variation and discriminatory value of biopsy features in inflammatory bowel disease. *Gut* 1994;**35**:961–8.
- 10 **Tanaka M**, Saito H, Fukuda S, *et al*. Simple mucosal biopsy criteria differentiating among Crohn disease, ulcerative colitis, and other forms of colitis: measurement of validity. *Scand J Gastroenterol* 2000;**35**:281–6.
- 11 **Tsang P**, Rotterdam H. Biopsy diagnosis of colitis: possibilities and pitfalls. *Am J Surg Pathol* 1999;**23**:423–30.
- 12 **Diamond J**, Anderson NH, Thompson D, *et al*. A computer-based training system for breast fine needle aspiration cytology. *J Pathol* 2002;**196**:113–21.
- 13 **Price AB**. Indeterminate colitis—broadening the perspective. *Curr Diagn Pathol* 1996;**3**:35–44.
- 14 **Jenkins D**, Bentley E, Fleming KA. Evidence-based cellular pathology: a systematic framework for pathological diagnosis and clinical decisions. *Curr Diagn Pathol* 2001;**7**:272–80.