**Appendices**

**Appendix 1. Guideline development process (BSG).**

|  |  |
| --- | --- |
| **Scope and purpose** |  |
| **Overall objective of the guideline** | To optimise the quality of histopathological reporting of biopsies taken for the diagnosis of inflammatory bowel disease ( IBD)To encourage consistency in histopathologists’ approach to IBD biopsy reportingTo encourage correlation between histopathological findings and clinical assessment |
| **The patient group covered** | Patients with IBD and relevant related gastrointestinal diseases, particularly those whose tissue is examined by a histopathologist. Mainly adults but also children. Male and female.  |
| **Target audience** | Histopathologists who report biopsies from patients with suspected or known IBD. Gastroenterologists and surgeons who are involved in the management of IBD. Hopefully, the document will be used to improve diagnostic accuracy and enhance clinical management.  |
| **Clinical areas covered** |  |
| **Stakeholder involvement** | Generally not relevant here. Document emphasises the importance of correlating histopathology with clinical findings and of good communication with clinical teams. |
| **Conflicts of interest** | None |
| **Rigour of development** |  |
| **Evidence Gathering** | **Search Strategy**Pub Med and other online search engines using keywords.Reference to previous sets of guideline documents from national and international societies and organisations (e.g., BSG, Royal Colleges, ECCO).Perusal of publications in electronic or paper form. **Inclusion Criteria*** Studies or reports dealing with adult or paediatric patients with IBD or related conditions.
* Most papers include histopathology.
* Peer reviewed papers or guidelines from established organisations or, where appropriate, material from textbooks.
* English language, at least in part.
* Evidence that the final diagnosis (e.g., inflammatory bowel disease, ulcerative colitis) was made using, at least, fairly robust criteria.

**Exclusion Criteria**Publication date before 1984 (with rare exceptions).**Search Terms**Multiple keywords in various combinations, e.g.* Pathology
* Histopathology
* Diagnosis
* Guidelines
* Discriminant features
* Inflammatory bowel disease
* Paediatric inflammatory bowel disease
* Ulcerative colitis
* Crohn’s disease
* Indeterminate colitis
* Activity
* Extent
* Reproducibility
* Interobserver variability
* Oesophageal / gastric / duodenal
* Diverticular colitis
* Diversion proctocolitis
* Infective colitis
* HIV
* Graft versus host disease
* Radiation colitis
* Ischaemic colitis
* Dysplasia

After the initial searches, further searches were based on additional relevant keywords. |
| **Review Process** | Publications printed out and/or saved electronically, and read, marked, and summarised by the author.Detailed tabulation of various findings from multiple papers, e.g., definitions of histological features; discriminant value of histological features in various situations; prevalence of histological features; reproducibility; etc.Attention to quality, relevance and nature of each study (e.g., initial biopsies only; initial plus non-initial biopsies; nature of comparisons; degree of follow up; multiple anatomical site vs. single site assessment; quality of analysis). |
| **Link between evidence and** **Recommendations** | As shown in tables and text. |
| **Piloting and peer review** | The author was asked in 2006 by the Chair of the Pathology Section committee of the BSG to revise the 1997 guidelines. Drafts were reviewed and amended by Pathology Section Committee of the BSG. They were also reviewed by representatives of the IBD Section of the BSG, by reviewers acting on behalf of the Clinical Services and Standards Committee of the BSG, and by representatives of the BSG Council. |

**Appendix 2. Implications**

 *Costs*

Implementation of the advice and recommendations in these guidelines might have the following impacts on costs and staff time in some institutions:

* more biopsies at endoscopy
* increased time spent by histopathologists reporting biopsies
* more work for technical staff, e.g., more biopsies, serial sections, deeper levels

The implementation of the recommendations might improve diagnostic accuracy. In turn, this might reduce the need for further biopsies while simultaneously helping to optimise medical management. This should reduce both workload and costs in the longer term.

*Resource availability*

In terms of organisation, the resources are available in most departments.

*Assessment of implementation*

After at least 18 months there should be an audit of the value of the guidelines, probably using email and a web-based survey. In particular, the ease of use and practical value of the suggested PAID reporting scheme should be analysed. The responses would help determine the content of a further update, which would be expected within the usual five year time frame.

**Competing interests:**

None

**Appendix 3. Grading quality of evidence & strength of recommendation.**

Oxford Centre for Evidence-based Medicine Levels of Evidence (March 2009) (for definitions of terms used see glossary at <http://www.cebm.net/?o=1116>)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Level**  |  **Therapy/Prevention,**  **Aetiology/Harm** | **Prognosis** | **Diagnosis** | **Differential diagnosis/symptom prevalence study** | **Economic and decision analyses** |
| 1a | SR (with homogeneity\*) of RCTs  | SR (with homogeneity\*) of inception cohort studies; CDR† validated in different populations | SR (with homogeneity\*) of Level 1 diagnostic studies; CDR† with 1b studies from different clinical centres | SR (with homogeneity\*) of prospective cohort studies  | SR (with homogeneity\*) of Level 1 economic studies |
| 1b | Individual RCT (with narrow Confidence Interval‡) | Individual inception cohort study with > 80% follow-up; CDR† validated in a single population | Validating\*\* cohort study with good††† reference standards; or CDR† tested within one clinical centre | Prospective cohort study with good follow-up\*\*\*\* | Analysis based on clinically sensible costs or alternatives; systematic review(s) of the evidence; and including multi-way sensitivity analyses |
| 1c | All or none§ | All or none case-series | Absolute SpPins and SnNouts†† | All or none case-series | Absolute better-value or worse-value analyses †††† |
| 2a | SR (with homogeneity\*) of cohort studies | SR (with homogeneity\*) of either retrospective cohort studies or untreated control groups in RCTs | SR (with homogeneity\*) of Level >2 diagnostic studies | SR (with homogeneity\*) of 2b and better studies | SR (with homogeneity\*) of Level >2 economic studies |
| 2b | Individual cohort study (including low quality RCT; e.g., <80% follow-up) | Retrospective cohort study or follow-up of untreated control patients in an RCT; Derivation of CDR† or validated on split-sample§§§ only | Exploratory\*\* cohort study with good††† reference standards; CDR† after derivation, or validated only on split-sample§§§ or databases | Retrospective cohort study, or poor follow-up | Analysis based on clinically sensible costs or alternatives; limited review(s) of the evidence, or single studies; and including multi-way sensitivity analyses |
| 2c | "Outcomes" Research; Ecological studies | "Outcomes" Research  |  | Ecological studies | Audit or outcomes research |
| 3a | SR (with homogeneity\*) of case-control studies |  | SR (with homogeneity\*) of 3b and better studies | SR (with homogeneity\*) of 3b and better studies | SR (with homogeneity\*) of 3b and better studies |
| 3b | Individual Case-Control Study |  | Non-consecutive study; or without consistently applied reference standards | Non-consecutive cohort study, or very limited population | Analysis based on limited alternatives or costs, poor quality estimates of data, but including sensitivity analyses incorporating clinically sensible variations. |
| 4 | Case-series (and poor quality cohort and case-control studies§§) | Case-series (and poor quality prognostic cohort studies\*\*\*) | Case-control study, poor or non-independent reference standard  | Case-series or superseded reference standards | Analysis with no sensitivity analysis |
| 5 | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" | Expert opinion without explicit critical appraisal, or based on physiology, bench research or "first principles" | Expert opinion without explicit critical appraisal, or based on economic theory or "first principles" |

**Grades of Recommendation**

|  |  |
| --- | --- |
| **A** | consistent level 1 studies  |
| **B** | consistent level 2 or 3 studies ***or*** extrapolations from level 1 studies |
| **C** | level 4 studies ***or*** extrapolations from level 2 or 3 studies  |
| **D** | level 5 evidence ***or*** troublingly inconsistent or inconclusive studies of any level |

*"Extrapolations" are where data is used in a situation that has potentially clinically important differences than the original study situation.*

*SR: Systematic review*

*RCT: Randomised control trial*

**Notes**

Users can add a minus-sign "-" to denote the level of that fails to provide a conclusive answer because:

* ***EITHER*** a single result with a wide Confidence Interval
* ***OR*** a Systematic Review with troublesome heterogeneity.

Such evidence is inconclusive, and therefore can only generate Grade D recommendations.

|  |  |
| --- | --- |
| \* | By homogeneity we mean a systematic review that is free of worrisome variations (heterogeneity) in the directions and degrees of results between individual studies. Not all systematic reviews with statistically significant heterogeneity need be worrisome, and not all worrisome heterogeneity need be statistically significant. As noted above, studies displaying worrisome heterogeneity should be tagged with a "-" at the end of their designated level. |
| "  | Clinical Decision Rule. (These are algorithms or scoring systems that lead to a prognostic estimation or a diagnostic category.) |
| "¡ | See note above for advice on how to understand, rate and use trials or other studies with wide confidence intervals. |
| § | Met when all patients died before the Rx became available, but some now survive on it; or when some patients died before the Rx became available, but none now die on it. |
| §§ | By poor quality cohort study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both exposed and non-exposed individuals and/or failed to identify or appropriately control known confounders and/or failed to carry out a sufficiently long and complete follow-up of patients. By poor quality case-control study we mean one that failed to clearly define comparison groups and/or failed to measure exposures and outcomes in the same (preferably blinded), objective way in both cases and controls and/or failed to identify or appropriately control known confounders. |
| §§§ | Split-sample validation is achieved by collecting all the information in a single tranche, then artificially dividing this into "derivation" and "validation" samples. |
| " "  | An "Absolute SpPin" is a diagnostic finding whose Specificity is so high that a Positive result rules-in the diagnosis. An "Absolute SnNout" is a diagnostic finding whose Sensitivity is so high that a Negative result rules-out the diagnosis. |
| "¡"¡ | Good, better, bad and worse refer to the comparisons between treatments in terms of their clinical risks and benefits. |
| " " "  | Good reference standards are independent of the test, and applied blindly or objectively to applied to all patients. Poor reference standards are haphazardly applied, but still independent of the test. Use of a non-independent reference standard (where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference') implies a level 4 study. |
| " " " "  | Better-value treatments are clearly as good but cheaper, or better at the same or reduced cost. Worse-value treatments are as good and more expensive, or worse and the equally or more expensive. |
| \*\* | Validating studies test the quality of a specific diagnostic test, based on prior evidence. An exploratory study collects information and trawls the data (e.g. using a regression analysis) to find which factors are 'significant'. |
| \*\*\* | By poor quality prognostic cohort study we mean one in which sampling was biased in favour of patients who already had the target outcome, or the measurement of outcomes was accomplished in <80% of study patients, or outcomes were determined in an unblinded, non-objective way, or there was no correction for confounding factors. |
| \*\*\*\* | Good follow-up in a differential diagnosis study is >80%, with adequate time for alternative diagnoses to emerge  |