J Clin Pathol 1994;47:1032–1034

Interobserver variation in the reporting of the histopathological grading of anal intraepithelial neoplasia

P S Carter, J P Sheffield, N Shepherd, D H Melcher, D Jenkins, P Ewings, I Talbot, J M A Northover

Abstract

***Aim*—To assess the consistency in the histological reporting of anal intraepithelial neoplasia (AIN) among experienced histopathologists.**

***Method*—One hundred anal biopsy specimens were retrieved from archival material at St Mark's Hospital, London and graded by five histopathologists according to criteria outlined by Fenger (six point scale, ranging from normal to invasive carcinoma).**

***Results*—There was only moderate agreement among the pathologists, with unweighted κ scores ranging from 0·09 to 0·48, and weighted κ scores of 0·17 to 0·60.**

***Conclusions*—There is considerable interobserver variation in the reporting of AIN. A simplified system of grading may help to abolish this.**

(J Clin Pathol 1994;47:1032–1034)

Colorectal Cancer Unit, St Mark's Hospital, London
P S Carter
I Talbot
J M A Northover

Department of Histopathology, Yeovil Hospital, Yeovil, Somerset
J P Sheffield

Department of Histopathology, Gloucester Royal Hospital, Gloucester
N Shepherd

Department of Histopathology, Royal Sussex County Hospital, Brighton
D H Melcher

Department of Histopathology, Queen's Medical Centre, Nottingham
D Jenkins

Somerset Health Authority, Taunton
P Ewings

Correspondence to:
Mr P S Carter, Department of Surgery, Basingstoke District Hospital, Basingstoke, Hants RG24 9NA

Accepted for publication 19 April 1994

Anal intraepithelial neoplasia (AIN) is a potentially premalignant change found in the anal canal. It was first described by the Danish pathologists Fenger and Nielsen.[1] The description and grading of the morphological changes, seen in both cervical and AIN, histologically, are very similar, differing only in the degree of keratinisation. Richart originally described the grading system in the cervix,[2] and these criteria have been clearly set out in a paper by Buckley *et al.*[3]

Fenger has used a similar system for reporting changes seen in the anus.[4] Both grading systems divide the severity of dysplasia into three grades: I, II, and III. The grades are differentiated by the amount of the epithelium that is affected. In grade I only the lowest third of the epithelium is affected, in grade II the lower two thirds are affected, and in grade III the whole epithelium is affected. In dysplasia the hyperchromatic nuclei are enlarged, with a relative increase in the nuclear to cytoplasmic volume ratio, mitotic figures, and acanthosis.

The histological assessment of cervical dysplasia seems to be subjective and prone to both inter- and intraobserver variation.[5] Robertson *et al* assessed the observer variability in the histopathological reporting of sections from 100 cervical biopsy specimens, by 12 consultant histopathologists.[6] This study showed poor differentiation between low grades of dysplasia but good agreement on the diagnosis of high grade dysplasia (CIN III). Very little agreement was found between observers on the identification of the changes of human papillovirus (HPV) infection.[6]

A second study by De Vet *et al* examined the same problem of observer variability. Four experienced histopathologists were asked to grade 106 cervical biopsy specimens into five groups according to the degree of dysplasia. This study showed again that there was considerable variability in the reporting of the same slides by different pathologists.[7]

The difficulties surrounding the diagnosis of dysplasia may produce problems when deciding on a therapeutic strategy. In the cervix there is good evidence to confirm that about 30% of grade III lesions will progress to invasive cancer within 10 years, if left untreated.[8] The natural history of the lower grades of cervical cancer is less well demarcated, although some studies have indicated a high progression rate of low grade lesions to high grade lesions over a short period of time. It is therefore unlikely that any large study of the natural history of low grade dysplasia will now be undertaken, due to ethical considerations.

This leaves the clinician with two choices, given the present grading system:

(1) The first choice is to treat all patients with repeated low grade smears aggressively, in an attempt to eradicate the disease. This often means, however, that the patients have frequent and multiple treatments for what may turn out to be a condition which will usually remain static or regress.

(2) The other option is to enrol the patients with repeated low grade smears into a regular colposcopic follow up clinic and only ablate if there is evidence of progression to more severe disease.

This therapeutic dilemma, which is partly related to grading difficulties, has been recognised by Richart. He has proposed an alternative grading system that divides dysplasia into two groups—major and minor.[9] This may allow clinicians to adopt an "expectant" policy with regard to the minor abnormalities, and confine surgery to the major dysplasia group.

The grading system described by Fenger for use in the anus has yet to be tested for its reproducibility.[4] The clinical decisions regarding treatment of AIN are different from those in the cervix. The natural history of AIN is yet to be defined. If it is assumed that high grade dysplasia in the anus will progress to invasive cancer in at least a small percentage of cases

then this group is likely to need ablative treatment. The problems arise with the types of patients who are at greatest risk of developing AIN.[10 11] These patients are immunosuppressed and will often heal poorly after surgery. Therefore, it is important to be able to restrict the number of patients needing to undergo ablative treatment to those most likely to develop invasive cancer in a short period of time. Subsequently, it is imperative to have a reliable, accurate, and reproducible grading system for anal pre-malignant changes.

Methods

One hundred histology slides were selected from archival material held in the pathology department of St Mark's Hospital. The slides were all of biopsy material from the anal canal, some were paraffin wax sections and some were fresh frozen sections. All slides had previously been reported (not by any of the participants in the study) as showing either normal anal mucosa, changes consistent with HPV infection, AIN (all grades), invasive anal cancer, or even a combination of the above.

All identification markings on the slides were removed and each slide was allocated a random number between 1 and 100 (a complete record was kept of the original identification data and the new numbers, to allow relabelling at the completion of the study).

The slides were sent to five histopathologists for reporting. All five histopathologists were consultants. Three consultants had a special interest in gastroenterological pathology and had considerable experience of assessing anal dysplasia. The other two consultant pathologists both had a special interest in cervical histopathology and had extensive experience of Richart's old system for assessing cervical intraepithelial dysplasia.

The results were assessed using both weighted and unweighted κ analyses. This statistical method allows pairs of data to be compared. The unweighted analysis is a direct comparison between each pair of pathologists' grading of each of the slides, irrespective of how far apart the grading is. The weighted κ analysis takes into account the ordinal nature of the scores and gives differential weights to differences in scores, according to how far apart the pathologists' scores are. The guidelines for the interpretation of κ scores are shown in table 1.[12]

For the κ analysis, a number was allocated to each grade: normal mucosa 0; inflammation/HPV infection 1; AIN I 2; AIN II 3; AIN III 4; squamous carcinoma 5; inadequate no grade.

Results

Overall, 495 diagnoses were made: five slides were reported as inadequate. The following levels of agreement occurred between the pathologists on examination of the 100 slides. Number of times: one single value recorded by all five = 12; exactly two different values recorded = 39; exactly three different values recorded = 39; exactly four different values recorded = 10; exactly five different values recorded = 0.

Table 2 shows the distribution of the grades for each of the pathologists.

The means shown above exclude the inadequate specimens. It is apparent from table 2 that pathologist B's scores were lower than average and pathologist C's were higher than average.

An example of a typical pairing is shown in table 3, which compares the results of pathologist A with those of pathologist E. Kappa scores for pathologist A compared with pathologist E (inadequates excluded): unweighted $\kappa = 0{\cdot}26$; weighted $\kappa = 0{\cdot}44$.

The unweighted κ scores for each pair of pathologists are summarised in table 4. The results of the unweighted κ scores ranged from 0·09 (poor agreement) to 0·48 (moderate agreement).

The weighted κ scores for each pair of pathologists are shown in table 5. The table of weighted κ scores shown above does, as expected, show improved figures, with scores ranging from 0·17 to 0·60.

Table 1 Intepretation of κ scores

Kappa score	*Interpretation*
< 0·20	Poor agreement
0·21–0·40	Fair
0·41–0·60	Moderate
0·61–0·80	Good
0·81–1·00	Very good

Table 2 Frequency distribution of scores by pathologist

	Pathologist					
Grade	*A*	*B*	*C*	*D*	*E*	*All*
0	6	21	1	22	18	68
1	34	52	15	34	35	170
2	25	6	25	10	11	77
3	16	9	25	11	5	66
4	12	7	28	18	25	90
5	5	5	4	5	5	24
Inadequate	2	0	2	0	1	5
Mean	2·1	1·4	2·8	1·8	2·0	2·0

Table 3 Example of a pairing: pathologist A compared with pathologist E

	Pathologist E						
Pathologist A	*0*	*1*	*2*	*3*	*4*	*5*	*Total*
0	4	2	0	0	0	0	6
1	6	18	3	1	5	1	34
2	4	12	4	1	4	0	25
3	1	3	2	2	8	0	16
4	1	0	2	1	8	0	12
5	0	0	0	0	0	4	4
Total	16	35	11	5	25	5	97

Kappa scores for pathologist A compared with pathologist E (inadequates excluded): unweighted $\kappa = 0{\cdot}26$; weighted $\kappa = 0{\cdot}44$.

Table 4 Summary of unweighted κ scores

	A	*B*	*C*	*D*	*E*
A	—	0·19	0·27	0·23	0·26
B	0·19	—	0·09	0·40	0·34
C	0·27	0·09	—	0·25	0·26
D	0·23	0·40	0·25	—	0·48
E	0·26	0·34	0·26	0·48	—

Table 5 Summary of weighted κ scores

	A	*B*	*C*	*D*	*E*
A	—	0·40	0·40	0·53	0·44
B	0·40	—	0·17	0·55	0·50
C	0·40	0·17	—	0·40	0·36
D	0·52	0·55	0·40	—	0·60
E	0·44	0·50	0·36	0·60	—

Discussion

Only moderate levels of agreement were reached even in the weighted κ analyses and although there was generally good agreement for invasive cancers and normal tissue, there was only poor agreement for all AIN grades.

Histopathological diagnosis and grading are deemed to be the "gold standard" in the diagnosis of many conditions. The introduction of quality control programmes in histopathology may highlight the differences between reports from different pathologists.[13]

The clinician's management is based on a knowledge of the natural history of a particular disease and the histopathological diagnosis of the disease. It is therefore important that the diagnosis and the grading are consistent. In the case of AIN the natural history is, as yet, unknown but if it follows a similar pattern to that of cervical intraepithelial neoplasia (CIN) then a third of high grade lesions will progress to invasive cancer if left untreated. *At present, despite the differences in the progressive potential of the various grades of CIN, all grades may receive some form of ablation. This may represent a degree of overtreatment.*

Many of the patients at risk of AIN are immunosuppressed and it is important in these patients to limit treatment to those who are at greatest risk of progression to invasive cancer. The present grading system is inadequate for a reliable differentiation to be made between the grades of AIN. This may be due partly to the effects of HPV infection, which can produce considerable interpretative difficulties.

In conclusion, it is clear from these studies that there is considerable interobserver variation in the reporting of AIN. A further study is now in progress to assess the reliability of a simplified classification, dividing AIN into two grades, high and low.

1 Fenger C, Thue Nielsen V. Dysplastic changes in the anal canal epithelium in minor surgical specimens. *Acta Pathol Microbiol Immunol Scand* 1981;**89**:463–5.
2 Richart R. Natural history of cervical intraepithelial neoplasia. *Clin Obstet Gynecol* 1967;**10**:748–84.
3 Buckley C, Butler E, Fox H. Cervical intraepithelial neoplasia. *J Clin Pathol* 1982;**35**:1–13.
4 Fenger C, Thue Nielsen V. Intraepithelial neoplasia in the anal canal. *Acta Pathol Microbiol Immunol Scand (Sect A)* 1986;**94**:343–9.
5 Cocker J, Fox H, Langley F. Consistency in the histological diagnosis of epithelial abnormalities of the cervix uteri. *J Clin Pathol* 1968;**21**:143–65.
6 Robertson A, Anderson J, Swanson Beck J, Burnett RA, Howatson SR, Lee FD, *et al.* Observer variability in histopathological reporting of cervical biopsy specimens. *J Clin Pathol* 1989;**42**:231–8.
7 De Vet H, Knipschild P, Schouten H, Koudstaal J, Kwee W-S, Willebrand D, *et al.* Interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol* 1990;**43**:1395–8.
8 McIndoe W, McLean M, Jones R, Mullins P. The invasive potential of carcinoma in situ of the cervix. *Obstet Gynecol* 1984;**64**:451–8.
9 Richart R. A modified terminology for cervical intraepithelial neoplasia. *Obstet Gynecol* 1990;**75**:131–3.
10 Scholefield J, Sonnex C, Talbot I, Palmer JG, Whatrup C, Mindel A, *et al.* Anal and cervical intraepithelial neoplasia: possible parallel. *Lancet* 1989;**ii**:765–9.
11 Carter P, de Ruiter A, Mindel A, Northover J. Anal intraepithelial neoplasia in HIV seropositive men. *Br J Surg* 1991;**78**:744.
12 Altman D. *Practical statistics for medical research.* London: Chapman and Hall, 1991.
13 Lee F, Burnett R. Quality assurance in histopathology. *J Pathol* 1987;**152**:247–51.