

The journal articles referenced below discuss the inter or intra-observational incongruities in pathologic studies concerning various states of neoplasia. A number of these studies concerning this observed variability utilize Kappa values. The Kappa value is the ratio of the proportion of times pathologists agree on a diagnosis compared to the proportion of times they could agree. If Kappa = 1, then there is perfect agreement. If Kappa = 0, then there is no agreement. The higher the value of Kappa, the stronger the agreement. In general, a high level of agreement occurs when Kappa values are above 0.5. Agreement is poor when kappa values are less than 0.3.

Interobserver and intraobserver bias exists in the interpretation of anal dysplasia, Colquhoun P, Nogueras JJ, Dipasquale B, Petras R, Wexner SD, Woodhouse S. Dis Colon Rectum. 2003 Oct;46(10):1332-6; discussion 1336-8.

Within this study the authors emphasize the fact that the natural history of progression from anal intraepithelial neoplasia to invasive carcinoma remains unproven. However, the risk of progression may be linked to the severity of dysplasia. As a result the authors feel that important therapeutic decisions are based on the severity of anal intraepithelial neoplasia. Nonetheless, consistency and reliability in the interpretation of anal intraepithelial neoplasia are unproven.

As part of this study, 190 anal biopsy specimens were identified for review of dysplasia with a six-point grade system from normal to invasive cancer, evidence of human papillomavirus infection, and quality of histology by three pathologists. The results indicated **poor to moderate agreement on grading of quality of histology** (weighted kappa score, **0.07-0.22**), human papillomavirus status (weighted kappa score, **0.24-0.53**), and dysplasia (weighted kappa score, **0.38-0.7**).

In fact, (and here is the important part) complete agreement between the original pathology and the three pathologists was observed in only **32 percent of cases**. Analysis of 86 slides previously read by one of the pathologists revealed only moderate agreement, with a weighted kappa score of 0.64.

The conclusion of this study was that significant interobserver and intraobserver bias exists in the interpretation of anal intraepithelial neoplasia. These inconsistencies may explain the uncertainty about the natural progression of anal intraepithelial neoplasia and the **varied results of surgery reported** for anal intraepithelial neoplasia in the literature.

Interobserver variation in the reporting of the histopathological grading of anal intraepithelial neoplasia, Carter PS, Sheffield JP, Shepherd N, Melcher DH, Jenkins D, Ewings P, Talbot I, Northover JM., J Clin Pathol. 1994 Nov;47(11):1032-4.

The intent of this study was to assess the consistency in the histological reporting of anal intraepithelial neoplasia (AIN) among experienced histopathologists. 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). It is interesting to note that there was **only moderate agreement** among the pathologists, with unweighted k scores ranging from **0.09 to 0.48**, and weighted k scores of **0.17 to 0.60**. The conclusion of the study was that there is considerable interobserver variation in the reporting of AIN.

The difficulty in histological interpretation of intraepithelial neoplasia is not limited to the area of AIN. It has also been observed in CIN.

Reporting cervical intra-epithelial neoplasia (CIN): intra- and interpathologist variation and factors associated with disagreement, Ismail SM, Colclough AB, Dinnen JS, Eakins D, Evans DM, Gradwell E, O'Sullivan JP, Summerell JM, Newcombe R. Histopathology. 1990 Apr;16(4):371-6.

In this particular study, eight histopathologists, based at different hospitals, who had previously examined 100 consecutive colposcopic cervical biopsies were circulated with the results of the initial study. The slides were then **'reblinded' and re-examined by the pathologists** who, as before, assigned them into one of six diagnostic categories.

The degree of interpathologist agreement for the seven observers who returned usable responses was characterized by kappa statistics and compared to the corresponding figures for the same observers from the previous study. Although some of the observers showed significant alterations in their diagnostic practices **there was persistent poor agreement for CIN 1 and 2, mediocre agreement for CIN 3 and excellent agreement for invasive carcinoma.**

Intra-observer agreement was consistently better than inter-observer agreement for each of the diagnostic categories. Significant differences were found among observers in the degree of intra-observer variability.

Inter- and intra-observer variation in the histopathological reporting of cervical squamous intraepithelial lesions using a modified Bethesda grading system, McCluggage WG, Walsh MY, Thornton CM, Hamilton PW, Date A, Caughley LM, Bharucha H. Br J Obstet Gynaecol. 1998 Feb;105(2):206-10.

The intent of this study was to assess inter- and intra-observer variation in the histopathological reporting of cervical colposcopic biopsies using a histologic

modification of the cytological Bethesda grading system and to determine the histologic profile of those cases, which resulted in diagnostic disagreement.

In this study, consecutive cervical 125 colposcopic biopsies were assessed independently by six experienced histopathologists. Cases were classified as normal, low grade squamous intraepithelial lesion or high grade squamous intraepithelial lesion. Six months later the process was repeated.

The degree of inter and intra-observer variation was assessed by kappa statistics. All cases in which there was less than perfect inter and intra-observer agreement were reviewed by the coordinator of the study.

In the first round of the study **inter-observer agreement was generally poor**, with unweighted and weighted kappa values ranging from 0.15 to 0.58 (**average 0.30**) and from 0.21 to 0.61 (**average 0.36**) respectively.

In this study there were marked difficulties in the separation of normal squamous epithelium from low grade squamous intraepithelial lesion and in the **separation of low grade from high grade squamous intraepithelial lesions**.

Histopathological review revealed that many of the difficulties in the separation of normal and low grade squamous intraepithelial lesion were in the distinction between superficial vacuolated cells and true koilocytes. In some cases, there was a full spectrum of diagnoses from normal to high grade squamous intraepithelial lesion.

The results of this study indicated that most pairs of pathologists can achieve fair inter-observer agreement in the reporting of cervical colposcopic biopsies using a

modified Bethesda system. Intra-observer agreement is also generally fair to good using this system.

It was suggested that a two tier grading system may be more appropriate for the histopathological reporting of these biopsies than the traditional three-tier intraepithelial neoplasia (CIN) system.

Inter-observer variation in histopathological diagnosis and grading of vulvar intraepithelial neoplasia: results of an European collaborative study, Preti M, Mezzetti M, Robertson C, Sideri M., BJOG. 2000 May;107(5):594-9.

This Italian study evaluated inter-observer the variability of vulvar intraepithelial neoplasia diagnosis and grading systems. The prospective study utilized histological sections of 66 vulvar biopsies. Six consultant pathologists working at different European institutions independently reviewed these vulvar biopsies.

The study addressed specimen adequacy, gross categorisation into benign or **neoplastic changes**, presence of atypical cytological pattern, presence of neoplastic architectural pattern, grade of vulvar intraepithelial neoplasia, presence of histopathologic associated findings for human papillomavirus infection.

The study utilized Kappa statistics where the frequency and the degree of disagreement were calculated by a symmetrical agreement matrix showing the number paired classifications.

A good agreement (overall weighted kappa = 0.65, unweighted kappa = 0.46) was observed for grading vulvar intraepithelial neoplasia. **Human papillomavirus infection associated findings and specimen adequacy** were the variables with less inter-observer agreement (overall weighted kappa **0.26 and 0.22**, respectively).

Exact agreement between two pathologists for grade of vulvar intraepithelial neoplasia was observed in only **63.6% of paired readings**. However, the rate of paired agreement reached 73.9% when considering vulvar intraepithelial neoplasia 2 and 3 as a single class. Conversely, only 5.0% of vulvar intraepithelial neoplasia 1 diagnoses were concordant in paired analysis..