

The contribution of MIB 1 in the accurate grading of vulvar intraepithelial neoplasia

M van Beurden, A J M de Craen, H C W de Vet, J L G Blaauwgeers, P Drillenburgh, M P W Gallee, N W de Kraker, F B Lammes, F J W ten Kate

Abstract

Aim—To determine the interobserver variation in scoring presence and grade of vulvar intraepithelial neoplasia (VIN) in haematoxylin/eosin (H/E) slides, MIB 1 slides, and the combined use of H/E and MIB 1 slides.

Methods—10 slides were stained with H/E and MIB 1 with each of the following diagnoses: normal vulvar skin, VIN 1, VIN 2, and VIN 3. Six observers first scored the H/E slides separately from the MIB 1 slides and second the combined H/E and MIB 1 slides.

Results—Unweighted group κ for MIB 1 was 0.62 and the weighted group κ was 0.91. This was significantly better than the unweighted group κ for H/E slides (0.47, $p = 0.023$) as well as the weighted group κ for H/E slides (0.82, $p = 0.014$). There was no improvement by the combined use of H/E and MIB 1 slides. VIN 2 is far less confused with VIN 3 in the combined use of H/E and MIB 1 slides (9%) than in H/E slides (38%) ($p = 0.007$). There is a tendency to grade VIN in a two tailed grading system rather than a three tailed grading system, which became more apparent with the combined use of H/E and MIB 1 slides.

Conclusions—The interobserver variation with sole use of MIB 1 is better than with the use of H/E stain in VIN. The use of MIB 1 in grading VIN diminishes confusion between VIN 2 and VIN 3 fourfold. A two tailed grading system for VIN seems already to work in daily practice.

(J Clin Pathol 1999;52:820-824)

Keywords: kappa test; MIB 1; vulvar intraepithelial neoplasia

Department of Pathology, Academic Medical Centre, Meibergdreef 9, 1105 AZ Amsterdam, The Netherlands

M van Beurden
A J M de Craen
H C W de Vet
J L G Blaauwgeers
P Drillenburgh
M P W Gallee
N W de Kraker
F B Lammes
F J W ten Kate

Correspondence to:
Dr van Beurden
email: beurden@nki.nl

Accepted for publication
17 June 1999

removal.¹ If one chooses to excise all visible lesions, it is therefore important to know which lesions do show VIN 3 and which do not. In this way one can leave as much vulvar skin as possible, avoiding psychological and sexual side effects from extensive surgery. It is not known how pathologists differ in their interpretation of VIN. However, they do differ in their interpretation of cervical intraepithelial neoplasia (CIN).^{3,4}

Measurement of cell proliferation may provide useful information on diagnosis and tumour prognosis. The Ki-67 monoclonal antibody is currently used in evaluating cellular proliferation rates of malignant tumours.⁵ A formalin resistant epitope of Ki-67 cell proliferation associated antigen is immunohistochemically detected by the MIB 1 monoclonal antibody.⁶ This has been proven to be the best proliferation marker for routine use in formalin fixed and paraffin embedded tissue sections.⁷ In preinvasive and invasive squamous neoplasms of the uterine cervix, the number of MIB1^{8,9} or Ki-67 positive cells increases with the severity of the lesion.¹⁰ Ki-67 expression has been described in normal vulvar epithelium and VIN 3,¹¹ but its value has never been examined systematically in the grading of VIN. VIN can have a spectrum of pathological changes, such as nuclear pleomorphism, hyperchromasia, altered epithelial maturation, cellular crowding, loss of normal keratinocyte polarity, and atypical mitotic figures.¹² VIN can be subclassified as VIN 1, 2, or 3, based on the extent of cellular disarray of the epithelium.¹² MIB 1 immunoquantitation is assessed by determining the percentage of labelled nuclei in the total population of nuclei and might therefore be a more accurate test than haematoxylin/eosin (H/E) staining for grading VIN.

In this study we sought to determine the interobserver variation in scoring the presence and grade of VIN in H/E slides, MIB 1 slides, and the combined use of H/E and MIB 1 slides.

Methods

Ten cases each of normal vulvar skin, VIN 1, VIN 2, and VIN 3 were retrieved from the pathology files of the department of pathology of the Academic Medical Centre. From each paraffin block two additional sections were recut. These sections were stained with H/E and MIB 1 (Immunotech, Coulter). The slides contained normal vulvar skin and VIN lesions. For MIB 1 antigen enhancement and optimisation of immunohistochemistry, slides were immersed in a 0.01 M sodium citrate dehydrated solution (pH 6.0) and then boiled for 10

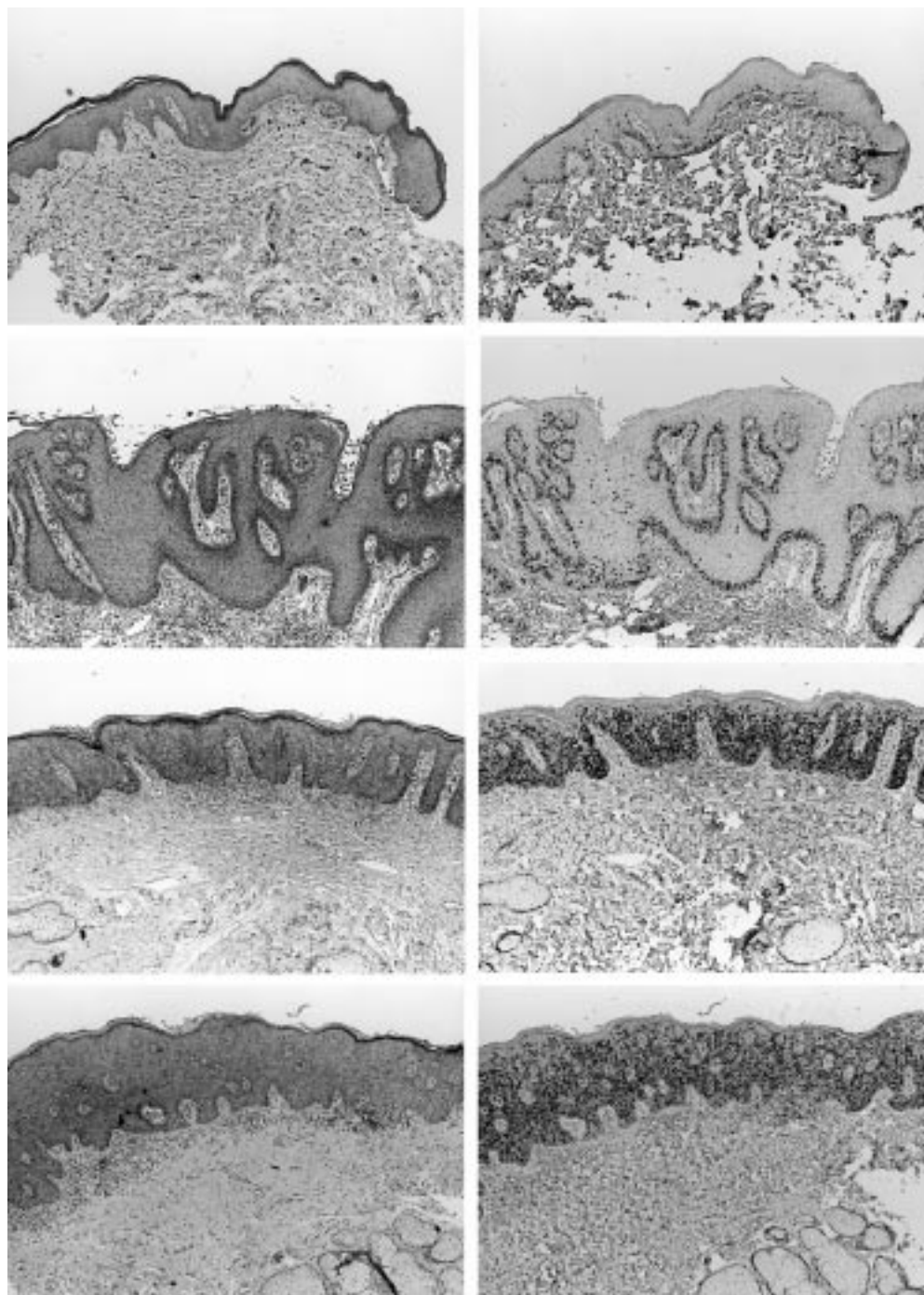


Figure 1 On the left from top to bottom: normal vulvar skin, VIN I, VIN II, VIN III. On the right from top to bottom: positive nuclei present in the basal two cell layers, the lower one third, the lower two thirds, and more than the lower two thirds of the epithelium.

minutes on a hot plate. After cooling overnight, the avidin-biotin complex methodology was used with diaminobenzidine as the chromagen. Nuclei were counterstained with haematoxylin. Five experienced pathologists and one gynaecological oncologist with special interest in gynaecological pathology each examined the 80 slides. They knew that their results would be compared, but they did not have any discussion over the grading criteria beforehand. Written definitions were given. First, the examiners scored the H/E and MIB 1 slides from each case separately, so that scoring of the H/E slides was not influenced by the corresponding

MIB 1 slides and vice versa. On the H/E slides, the presence and grade of VIN (1–3) was scored according to the following definitions¹²:

- Normal vulvar skin: no distorted architecture;
- VIN 1: cellular disarray involving the lower one third of the epithelium;
- VIN 2: cellular disarray involving the lower two thirds of the epithelium;
- VIN 3: cellular disarray involving more than the lower two thirds of the epithelium.

On the MIB 1 slides, the extent to which nuclei were positive throughout the epithelium was scored, depending on whether positive

Table 1 Frequency distribution of the grading results among six observers on haematoxylin and eosin slides and HE-MIB 1 slides (*italic*)

Observer	Vulvar epithelium							
	Normal (%)		VIN 1 (%)		VIN 2 (%)		VIN 3 (%)	
1	35	18	13	15	23	33	30	35
2	18	28	38	20	20	30	25	23
3	20	21	30	26	20	18	30	36
4	33	35	15	13	18	20	35	33
5	10	13	30	35	23	18	38	35
6	20	23	30	25	18	30	33	23
Average	23	23	26	22	20	30	32	31

Table 2 Frequency distribution of the gradings of the six observers on MIB 1 slides

Observer	Positive nuclei in vulvar epithelium			
	Basal 2 cell layers (%)	Lower one third (%)	Lower two thirds (%)	> Lower two thirds (%)
1	45	5	38	13
2	43	5	15	38
3	33	15	25	28
4	40	5	10	45
5	45	13	15	28
6	43	5	23	30
Average	42	8	21	30

nuclei were present in either the basal two cell layers, the lower one third, the lower two thirds, or more than the lower two thirds of the epithelium. Outside the scored category for MIB 1, approximately 5% of positive nuclei in the remainder of the epithelium were ignored (fig 1). At a second microscopic examination two months later, the H/E and MIB 1 slides from each case were scored together, so this time assessment of the H/E slides was meant to be influenced by the MIB 1 slides and vice versa. This time, only the presence and grade of VIN (1–3) was scored.

STATISTICS

The interobserver variation was established by determining the κ value.¹³ In addition to unweighted κ values, weighted κ values were calculated.¹⁴ In the weighted κ value, agreement is given more weight if observers differ by only one category than if they disagree more than one category. A group κ can be established in addition to calculating κ values between pairs of observers. In the group κ , the average observed agreement is compared to the average agreement made by chance over all pairs of observers and over all slides. Both an

unweighted and a weighted group κ were calculated.¹⁵ We assumed that differences between group κ values had a normal distribution, hence the significance of differences was assessed by calculation of z values.

Results

Of 720 observations (40 H/E slides, 40 MIB 1 slides, 40 H/E–MIB 1 slides, six observers), 719 were recorded, as observer 3 was unable to assign one score in a combined reading of H/E–MIB 1 slides. Tables 1 and 2 show the frequency distributions of the scores of the different categories for H/E slides, H/E–MIB 1 slides, and MIB 1 slides. In the MIB 1 slides, the presence of positive nuclei in the lower one third of the epithelium was scored less (8%) than in the other categories. This difference disappeared if the presence of positive nuclei in the basal two cell layers and the lower one third of the epithelium were considered together as one category (50%) and compared to the presence of positive nuclei in the lower two thirds and more than the lower two thirds of the epithelium, taken together as one category.

To compare the agreement between the individual observers for H/E slides, MIB 1 slides, and H/E–MIB 1 slides, unweighted and weighted κ values for each pair of observers were calculated (table 3). The unweighted κ values ranged from 0.22 to 0.63 for H/E slides, from 0.38 to 0.77 for MIB 1 slides, and from 0.41 to 0.77 for H/E–MIB 1 slides. The weighted κ values ranged from 0.73 to 0.88 for H/E slides, from 0.85 to 0.95 for MIB 1 slides, and from 0.79 to 0.94 for H/E–MIB 1 slides. The κ values were not better between pathologists than between pathologists and the gynaecological oncologist, nor between pathologists within a single institute. The unweighted group κ was 0.47 for H/E slides, 0.62 for MIB 1 slides, and 0.60 for H/E–MIB 1 slides. The weighted group κ was 0.82 for H/E slides, 0.91 for MIB 1 slides, and 0.87 for H/E–MIB 1 slides. This improvement in unweighted and weighted group κ values between H/E slides and MIB 1 slides was significant ($p = 0.023$ and 0.014 , respectively). However, the improvement in unweighted and weighted group κ values between H/E slides and H/E–MIB 1 slides was not significant ($p = 0.08$ and 0.26 , respectively).

In order to assess which in categories the H/E slides and the H/E–MIB 1 slides differed,

Table 3 Kappa values for agreement between pair of pathologists for haematoxylin and eosin (H/E) slides, MIB 1 slides (*bold*), and H/E–MIB 1 slides (*italic*) using four categories of classification

Observers	Unweighted kappa values																
	1	2	3	4	5	6	1	2	3	4	5	6					
Weighted κ values	1		0.51	0.54	<i>0.56</i>	0.54	0.51	<i>0.41</i>	0.55	0.38	<i>0.60</i>	0.22	0.61	<i>0.53</i>	0.44	0.60	<i>0.43</i>
	2	0.83	0.88	<i>0.82</i>		0.60	0.65	<i>0.49</i>	0.51	0.77	<i>0.77</i>	0.35	0.70	<i>0.54</i>	0.63	0.74	<i>0.67</i>
	3	0.80	0.88	<i>0.83</i>	0.86	0.92	<i>0.86</i>		0.53	0.61	<i>0.55</i>	0.39	0.55	<i>0.68</i>	0.46	0.76	<i>0.76</i>
	4	0.85	0.85	<i>0.79</i>	0.86	0.94	<i>0.94</i>	0.88	0.92	<i>0.89</i>		0.44	0.58	<i>0.58</i>	0.50	0.66	<i>0.64</i>
	5	0.73	0.91	<i>0.85</i>	0.76	0.92	<i>0.83</i>	0.78	0.90	<i>0.91</i>	0.79	0.88	<i>0.86</i>		0.42	0.64	<i>0.64</i>
	6	0.83	0.91	<i>0.83</i>	0.85	0.95	<i>0.90</i>	0.81	0.94	<i>0.93</i>	0.87	0.92	<i>0.90</i>	0.79	0.93	<i>0.88</i>	

Unweighted group κ (H/E) = 0.47; 95% confidence interval (CI) 0.38 to 0.56.

Weighted group κ (H/E) = 0.82; 95% CI 0.76 to 0.88.

Unweighted group κ (MIB 1) = 0.62; 95% CI 0.52 to 0.71.

Weighted group κ (MIB 1) = 0.91; 95% CI 0.87 to 0.95.

Unweighted group κ (H/E–MIB 1) = 0.60; 95% CI 0.48 to 0.71.

Weighted group κ (H/E–MIB 1) = 0.87; 95% CI 0.81 to 0.93.

Table 4 Frequency distribution (%) of the scores of five observers on 40 haematoxylin and eosin slides and 40 HE-MIB 1 slides (italic), conditional on the judgement of the other observer

Vulvar skin	Total number of scores (N)		Vulvar skin							
			Normal (%)		VIN 1 (%)		VIN 2 (%)		VIN 3 (%)	
Normal	54	54	47	69	44	31	9	0	0	0
VIN 1	62	53	40	48	56	52	4	0	0	0
VIN 2	48	59	4	9	22	40	36	43	38	9
VIN 3	76	73	0	0	3	0	18	23	78	77

the disagreements in grading are shown in table 4. The second column from the left (total number of scores) shows the combined frequency distribution of the scores of six observers for H/E slides and H/E-MIB 1 slides. The rows show the frequency distribution of the other five observers, when one observer assigned the score mentioned in the left hand column. For instance, for the slides that were scored as VIN 1 in the H/E slides by one observer, the distribution of the others were: 40% normal vulvar skin, 56% VIN 1, 4% VIN 2, and 0% VIN 3. Normal vulvar skin and VIN 1 were equally confused with each other (40–56%), but not with VIN 2 (4–9%) or VIN 3 (0%) in H/E slides. This sharp distinction between normal vulvar skin and VIN 1 on the one hand and VIN 2 and VIN 3 on the other hand appeared even more striking (though the difference was not significant) in the H/E-MIB 1 slides compared with the H/E slides, where no confusion was found at all. VIN 2 was hardly ever confused with normal vulvar skin, sometimes with VIN 1, but nearly twice as often with VIN 3 in H/E slides. VIN 2 is far less confused with VIN 3 in H/E-MIB 1 slides (9%) than in H/E slides (38%) (χ^2 test, $p = 0.007$).

Discussion

The κ statistic is the measure of choice for assessing interobserver variation. It corrects for chance agreement, which is not taken into account if percentages of agreement between observers are compared. Unweighted κ values depend on the number of categories one wants to distinguish: if detailed subdivisions are required, the task becomes more difficult and the κ values will be lower.¹³ This led to the development of the weighted κ ,¹⁴ which take into account the degree to which disagreement concerns adjacent or more remote categories; κ values below 0.40 may be taken to represent poor agreement, values between 0.40 and 0.75 fair to good agreement, and values above 0.75 excellent agreement.¹⁶ Unweighted⁴ and weighted⁵ kappa values have been calculated for grading CIN on routinely stained H/E sections. It has been shown that the group κ for grading CIN is improved significantly if observers agree beforehand on which morphological characteristics should be considered relevant for grading CIN.¹⁷ This can be optimised after a consensus meeting between observers through a joint session behind a discussion microscope about the method of grading CIN.¹⁸ This may explain in part why in this study the interobserver variation for scoring presence and grading of VIN in H/E slides was

already fair (unweighted group κ 0.47) to excellent (weighted group κ 0.82), as written definitions about grading were handed out at the scoring. Another explanation might be that the observers were all experienced in gynaecological pathology. Therefore, an improvement in unweighted and weighted group κ values between H/E slides and H/E-MIB slides is difficult to establish. However, there was a significant improvement in unweighted as well as in weighted group κ values between H/E slides and MIB 1 slides. This is not surprising, as only one variable, namely positive nuclei, had to be taken into account in the MIB 1 slides. By contrast, many characteristics play a role in determining the presence and grading of VIN in H/E slides, such as the extent of cellular disarray of the epithelium, nuclear pleomorphism, hyperchromasia, altered epithelial maturation, cellular crowding, loss of normal keratinocyte polarity, and atypical mitotic figures. We assumed that the presence and extent of positive nuclei in four categories in the MIB 1 slides correspond with normal vulvar skin and VIN 1–3. It is unknown at present how MIB 1 immunoquantitation in general and the categories used in this study for presence and grade of VIN are correlated with clinical outcome of VIN. Therefore, the use of MIB 1 on its own for assessing the presence and grading of VIN is unwarranted at present. However, MIB 1 staining should be evaluated in correlation studies in which cytometric analyses are also evaluated.

Positive nuclei present in the lower one third of the epithelium was scored significantly less in MIB 1 slides than in the other categories, although in H/E slides and H/E-MIB 1 slides such a strong difference was not found for VIN 1, compared with the other categories. This may be explained by the categories used for the presence of positive nuclei in MIB 1 slides in our study. The observers had to make a distinction between positive nuclei present in the basal two cell layers and the lower one third of the epithelium. In many VIN cases the basal two cell layers and the lower one third of the epithelium nearly enclose the same part of the epithelium.

It is clear that the combination of H/E and MIB 1 significantly diminishes the confusion between VIN 2 and VIN 3, by 38% in the H/E slides and by 9% in the H/E-MIB 1 slides, as can be seen in table 4. It also seems from table 4 that observers were more likely to grade the slides into two categories (normal vulvar skin and VIN 1 *v* VIN 2 and VIN 3) than into four categories. We showed on the one hand that normal vulvar skin and VIN 1 are easily confused, but hardly ever with VIN 2 and never with VIN 3. This becomes even clearer in the H/E-MIB 1 slides. However, VIN 2 is hardly ever confused with normal vulvar skin, sometimes with VIN 1, but nearly twice as often with VIN 3. VIN 3 is never confused with normal vulvar skin, hardly ever with VIN 1, and sometimes with VIN 2. This finding of a two tailed grading system for VIN may reflect the fact that CIN is graded as low grade squamous intraepithelial neoplasia (LGSIL) and high

grade squamous intraepithelial neoplasia (HGSIL). This could readily influence the grading of VIN.

CONCLUSIONS

Pathologists can use MIB 1 in grading VIN, as confusion of VIN 2 with VIN 3 is lowered fourfold. This could have implications for the management of patients with VIN 3, for whom standard treatment advice is to remove only VIN 3 lesions. It is worthwhile considering the implementation of a two tailed grading system—high grade and low grade squamous intraepithelial lesions—for VIN, as in daily practice this already seems to be what is done.

- 1 van Beurden M, van der Vange N, ten Kate FJW, et al. Restricted surgical management of vulvar intraepithelial neoplasia 3: focus on exclusion of invasion and on relief of symptoms. *Int J Gynecol Cancer* 1998;**8**:73–7.
- 2 Kaufman RH. Intraepithelial neoplasia of the vulva. *Gynecol Oncol* 1995;**56**:8–21.
- 3 Ismail SM, Colglough AB, Dinnen JS, et al. Observer variation in histopathological diagnosis and grading of cervical intraepithelial neoplasia. *BMJ* 1989;**298**:707–10.
- 4 Vet HCW, Knipschild PG, Schouten HJA, et al. Interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol* 1990;**43**:1395–8.
- 5 Gerdes J, Schwab U, Lemke H, et al. Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation. *Int J Cancer* 1983;**31**:13–20.
- 6 Cattoretti G, Becker MHG, Key G, et al. Monoclonal antibodies against recombinant parts of the Ki-67 antigen (MIB and MIB3) detect proliferating cells in microwave-processed formalin-fixed paraffin sections. *J Pathol* 1992;**168**:357–63.
- 7 Rose DSC, Maddox PH, Brown DC. Which proliferation markers for routine immunohistology? A comparison of five antibodies. *J Clin Pathol* 1994;**47**:1010–14.
- 8 Garzetti GG, Ciavattini A, De Nictolis M, et al. MIB 1 immunostaining in cervical intraepithelial neoplasia: prognostic significance in mild and moderate lesions. *Gynecol Obstet Invest* 1996;**42**:261–6.
- 9 Bulten J, van der Laak JA, Gemmink JH, et al. MIB1, a promising marker for the classification of cervical intraepithelial neoplasia. *J Pathol* 1996;**178**:268–73.
- 10 Al-Saleh W, Delvenne P, Greimers R, et al. Assessment of Ki-67 antigen immunostaining in squamous intraepithelial lesions of the uterine cervix. Correlation with the histologic grade and human papillomavirus type. *Am J Clin Pathol* 1995;**104**:154–60.
- 11 Van Hoesven KH, Kovatich AJ. Immunohistochemical staining for proliferating cell nuclear antigen, BCL2, and Ki-67 in vulvar tissues. *Int J Gynecol Pathol* 1996;**15**:10–16.
- 12 Wilkinson EJ. Normal histology and nomenclature of the vulva, and malignant neoplasms, including VIN. *Dermatol Clin* 1992;**10**:283–96.
- 13 Cohen J. A coefficient of agreement for nominal scales. *Educ Psychol Meas* 1960;**20**:37–46.
- 14 Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull* 1968;**70**:213–30.
- 15 Schouten HJA. Nominal scale agreement among observers. *Psychometrika* 1986;**51**:453–66.
- 16 Fleiss JL. The measurement of interrater agreement. In: Fleiss JL, ed. *Statistical methods for rates and proportions*, 2nd ed. New York: John Wiley and Sons, 1981:212–36.
- 17 Vet HCW, Knipschild PG, Schouten HJA, et al. Sources of interobserver variation in histopathological grading of cervical dysplasia. *J Clin Epidemiol* 1992;**45**:758–90.
- 18 Vet HCW, Koudstaal J, Kwee WS, et al. Efforts to improve interobserver agreement in histopathological grading. *J Clin Epidemiol* 1995;**48**:869–73.