Multivariate discriminant analysis of normal, intraepithelial neoplasia and human papillomavirus infection of the uterine cervix samples

E. Artacho-Pérula[1], R. Roldán-Villalobos[1], J. Salas-Molina[2] and R. Vaamonde-Lemos[1]
[1]Department of Morphological Sciences and [2]Reina Sofía Hospital, University of Córdoba, Córdoba, Spain

Summary. The present investigation studies the role of multivariate statistical methods on quantitative histopathological features of cells in uterine cervix epithelium to discriminate between normal and abnormal uterine cervix samples. 143 histological specimens were included in the study involving normal cervix, cervical intraepithelial neoplasia (CIN) lesions and cervical human papillomavirus (HPV) infection with and without CIN (condyloma-CIN and condyloma-NCIN groups, respectively). Deep, middle and superficial regions of the cervical squamous epithelium were morphometrically analyzed. Identification of normal cervix from pathological cases was highly achieved with a specificity of 100%. The application of discriminant statistical method within pathological specimens showed an acceptable percentage of cases correctly classified; thus, an efficiency of 83.0% and 74.6% was obtained in order to discriminate within CIN and condyloma-CIN grades respectively. These percentages increased when differentiation between each grade of CIN versus condyloma-CIN were considered, using only 1-3 morphometrical parameters. Our findings indicate that the combination of nuclear and cytoplasmic quantitative features, specially size parameters, permit a high correct percentage classification of cervix samples. The discrimination process was better when few diagnostic categories were included; however, 100% specificity for normal samples was always reached.

Key words: Uterine cervix, Human papillomavirus, Cervical intraepithelial neoplasia, Discriminant analysis

Introduction

There is a considerable amount of information available on cervical human papillomavirus (HPV)

infections (Syrjänen, 1983; Gissmann et al., 1984; Kadish et al., 1986; Syrjänen et al., 1987) and cervical intraepithelial neoplasia (CIN) (Fu et al., 1989; Montag et al., 1989). Recently, morphometry, a technique used extensively in biological sciences, has quantitatively described the nuclear and cytoplasmic features of the intraepithelial cells of the uterine cervix for both control and pathological biopsies (Twiggs et al., 1981; Hartman et al., 1986; Hall et al., 1988; Bibbo et al., 1989; Mariuzzi et al., 1989).

Many authors have reported the quantitative detection of uterine cervix disorders (Abdul-Karim et al., 1982; Boon and Kok, 1985; Hall et al., 1988; Montag et al., 1989). Thus, measurements of nuclei and cytoplasm in intraepithelial cells have shown that the use of nuclear area as a diagnostic criterion to differentiate between normal and abnormal cervix is adequate (Tosi et al., 1988). However, isolated quantification of the uterine cervix cells are of little value for discrimination within pathological processes.

In the present study, the diagnostic value of nuclear and cytoplasmic size and shape in the detection of HPV infections and CIN lesions is assessed together with multivariate discriminant analysis to obtain a better differentiation of these processes. Thus, morphometric quantification of the deep, middle and superficial layers of stratified uterine cervix epithelium cells is performed and their features investigated in order to provide a better dignostic criterion.

Materials and methods

The study was performed on paraffin-embedded cervical tissue specimens from archival cases from the Reina Sofia Hospital (Córdoba, Spain). Sections from paraffin blocks were cut at 6 μm and stained with haematoxylin and eosin. All cases were reviewed and the diagnoses were confirmed independently by two pathologists. Several weeks later, the material was re-assessed by the same pathologists. Samples were excluded from the study when they showed autolitic and

deteriorated images, or when there was disagreement between the pathologists.

A total of 143 samples were used for further study, comprising the following groups: 1) Normal: 16 samples; 2) Cervical intraepithelial neoplasia (CIN) lesions: 55 samples (including 7, 11 and 37 samples for CIN I, II and III, respectively); 3) Cervical human papillomavirus (HPV) infection without CIN (condyloma-NCIN): 10 samples; and 4) HPV infection with CIN (condyloma-CIN): 62 samples (including 23, 29 and 10 samples for condyloma-CIN I, II and III, respectively). The classification of the CIN I, II and III was based on histopathological criteria: CIN I, mild dysplasia; CIN II, moderate dysplasia; and CIN III, severe dysplasia and carcinoma in situ (Abdul-Karim et al., 1982).

The quantitative analysis was performed by one observer experienced in the use of the graphic tablet (E.A-P). The Leitz A.S.M. semi-automatic image analyser equipped with a cursor was used for all measurements. The cursor is fitted with a light-emitting diode, which can be observed through a Leitz Dialux 20 microscope with a drawing tube for measurements of microscopic images on the graphic tablet. Calibration of the magnifiction of this microscope/drawing system was performed by using an object microscope graticule of equidistant (10 μm) lines. The contours of cytoplasm and nucleus were traced and the data was calculated by microcomputer and recorded in a personal computer.

For each diagnosis and sample, an average of six microscopic fields (range 3-6) were chosen at random within the epithelium where the whole thickness of the epithelium was seen. Within each field, stratified squamous epithelium cells were examined at a final magnification of x 2,200 at the three levels: deep, middle and superficial layers. Thirty cells per layer (range 25-40) were measured at random. This number of cells was selected according to studies by Tosi et al. (1986, 1988).

The morphometric parameters selected for analysis were the area, the perimeter, the maximal, minimal and equivalent circle diameters, nuclear/cytoplasmic ratio (N/C ratio) and several shape factors (Meijer et al., 1980; Marchevsky et al., 1987; Tosi et al., 1988) which included the following: a) Form factor PE ($4 \cdot \pi \cdot area/perimeter^2$), its value is 1.00 for a circle and <1.00 for ellipse and irregular structures; b) Form factor ELL (minimal diameter/maximal diameter), its value is 1.00 for a circle and <1.00 for elliptical structures; c) Form factor AR ($area/((\pi/4) \cdot maximal\ diameter \cdot minimal\ diameter)$), its value is 1.00 for a circle and an ellipse and <1.00 for irregular structures; d) Form factor CO ($perimeter/(area)^{1/2}$), this parameter takes the minimum value of 3.54 for a circular contour; and e) Form factor CI ($perimeter/(\pi \cdot equivalent\ circle\ diameter)$), its value is 1.00 for a perfect circle and >1.00 for irregular structures. The quantitative information was obtained for both nuclear and cytoplasmic profiles of the cells.

The mean, standard deviation and coefficient of variation of each feature were computed. Normal distribution was tested with the Kolmogorov-Smirnov test; an alpha level of <0.05 was accepted for significance.

The ability of morphometry to distinguish between different processes was investigated by multivariate discriminant analysis. A total of 63 features were included in each discriminant analysis (11 size parameters and 10 shape parameters in each cervix epithelium region). The 12 discriminant analyses performed are shown in Figure 1. Analyses 1-6 include normal group for discrimination with pathological processes. Analyses 7-12 allow discrimination within pathological processes. The efficiency, sensitivity and specificity was also calculated according to Collan (1989).

All statistical analyses were performed on a Mitac MPC 2000SL personal computer using the SPSS/PC Statistical Software Package procedure DSC. The DSC performs linear discriminant analysis for two or more groups. The aim of discriminant analysis is to classify cases into one of several mutually exclusive groups, based on their values for a set of predictor variables. The classification rule is developed on cases for which group membership is known. The rule can then be used to classify cases for which group membership is not known. A stepwise analysis was carried out where Wilks' method was used for entering variables into the analysis phase; at each step, the variable which minimizes the overall Wilks' lambda is entered. Thus, classification function coefficients (Fisher's linear discriminant functions), discriminant scores and classification information of the cases, and classification results table are displayed.

Results

Figure 1 summarizes the number of variables chosen by multivariate discriminant analyses and the accuracy of the classification for each preselected group. It is evident that the normal group could indeed be discriminated from pathological cases. The first to sixth discriminant analyses showed a 100% recognition rate of normal group samples. The features chosen were mainly size parameters.

The worst overall correct recognition rate appeared when eight cytopathological diagnoses were included (discriminant analysis number one -DA1-). In spite of the high number of variables selected, there was only efficiency discrimination in 48.55% of the cases. If the samples were divided into four groups (normal, CIN, condyloma-NCIN and condyloma-CIN) the total percent of diagnostic concordance was 61.76% (DA2). When the discrimination was performed between three groups (normal, CIN and HPV infection with/without CIN) the overall accuracy rate increased to 68.38% through six variables (five corresponding to size parameters and one to shape factor parameter) (DA3).

The power of discrimination between groups increased when the statistical analyses were performed

to determine which features separated normal group from CIN group (DA4), from condyloma-NCIN group (DA5) and from condyloma-CIN group (DA6). The overall accuracy rates were 89.86%, 87.50% and 85.54% respectively, while the specificity was always 100%.

These results were obtained with descriptors concerning nuclear size and nuclear/cytoplasmic ratios. The discrimination functions and classification results are shown in Table 1.

The 7th to 12th discriminant analyses were performed

	DA1	DA2	DA3	DA4	DA5	DA6	DA7	DA8	DA9	DA10	DA11	DA12
CIN I												
CIN II												
CIN III												
Normal												
Condyloma-NCIN												
Condyloma -CIN I												
Condyloma-CIN II												
Condyloma-CIN III												
Number of variables	8	8	6	2	2	1	8	8	7	1	3	3
Percent of cases correctly calssified	28.6	60.4	66.0	87.3	100.0	100.0	57.1	100.0	77.3	85.7	81.8	81.8
	54.5	100.0	100.0	100.0	70.0	82.6	40.0	90.0	74.1	90.9	85.2	85.2
	43.2	60.0	63.8				50.0	77.8	70.0			
	100.0	54.2					63.6					
	40.0						37.0					
	45.5						60.0					
	33.3											
	60.0											

Fig. 1. Diagnostic groups and discriminant analyses (DA) included in our study. Each box represents a preselected group. For instance, DA1 performs a discriminant analysis between eight groups whereas DA2 analyzes discrimination between four groups (CIN grades and condyloma-CIN grades are joined). The number of variables for each DA is shown. The percentage of cases correctly classified is printed in order for each preselected group; thus, DA2 classify correctly 60.4%, 100.0%, 60.0% and 64.2% of CIN, normal, condyloma-NCIN and condyloma-CIN cases, respectively.

Table 1. Classification matrix for DA4, DA5 and DA6, and coefficients of the discriminant functions.

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		NORMAL	CIN	
Normal	14	14	0	100.0%
CIN	55	7	48	87.3%
			Total % concordance: 89.86%	

DISCRIMINANT FUNCTIONS				
	NORMAL	CIN	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-33.89	-57.92		
NMaD-M	7.29	9.16	39.54	0.447
N/C ratio-S	63.69	95.06	39.14	0.624

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		NORMAL	CONDYLOMA-NCIN	
Normal	14	14	0	100.0%
Condyloma-NCIN	10	3	7	70.0%
			Total % concordance: 87.50%	

DISCRIMINANT FUNCTIONS				
	NORMAL	CONDYLOMA-NCIN	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-93.51	-126.27		
N/C ratio-B	272.27	313.82	13.90	0.430
NMiD-S	26.39	31.12	10.60	0.675

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		NORMAL	CONDYLOMA-CIN	
Normal	14	14	0	100.0%
Condyloma-CIN	69	12	57	82.6%
			Total % concordance: 85.54%	

DISCRIMINANT FUNCTIONS				
	NORMAL	CONDYLOMA-CIN	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-29.07	-44.87		
NEcD-S	14.39	17.95	40.49	0.677

NMaD-M: nuclear maximal diameter for intermediate cells; N/C ratio-S: nuclear/cytoplasmic ratio for superficial cells; N/C ratio-B: nuclear/cytoplasmic ratio for basal cell; NMiD-S: nuclear minimal diameter for superficial cells; NEcD-S: nuclear equivalent circle diameter for superficial cells.

to differentiate within pathological groups. The DA7 (considering six sets of patients -three CIN and three condyloma-CIN-) did not clarify these entities sufficiently, since the percentage of diagnostic concordance was only 50%.

In practice, it is interesting to discriminate within CIN group and within condyloma-CIN group. Eight and seven variables were selected respectively (DA8 and DA9). The discrimination was better when considering CIN stages (83.02% of the total cases) than within condyloma-CIN lesions (74.58%). In both analyses, the cases correctly classified decreased with the increase of tumoral dedifferentiation (Fig. 1).

Another decision was to make a correct classification between CIN and condyloma-CIN within the same stage of dedifferentiation. Thus, DA10 to DA12 were carried out (Fig. 1). The variables selected corresponded to both cell size and shape characteristics of superficial and middle layers of the cervix epithelium. The percentage of diagnostic concordance in relation to cytopathological classification was greater when the CIN 1 versus condyloma-CIN 1 were considered (89.66% of the total cases); this percentage decreased to 80.85% when CIN III versus condyloma-CIN III were computed. The discrimination functions and classification results are shown in Table 2.

Discussion

The present study has shown that multivariate statistical techniques are useful for the routine diagnosis and/or classification of normal and pathological samples. The nuclear and cytoplasmic features, both size and shape, are the most common parameters for application in Pathology (Boon and Kok, 1985; Baak et al., 1988; Fu et al., 1988; Norris et al., 1989). After a selection of the problem of interest, the application of an appropriate general procedure and measurements of the morphological characteristics of the samples (Artacho-Pérula et al., 1993), the researchers performed several multivariate analyses. The morphometrical data were analyzed to provide a powerful discriminating tool to determine which parameters were the most suitable to classify the individual cases into specific groups. Furthermore, discriminant analyses provided a probability statement for the morphometric classification of the cases.

Classification rules are seldomly 100% valid. This percentage is obtained with extreme difficulty (Baak et al., 1982). Whereas changes in tissues moving away from normality are more easily detected, the correct classification within pathological entities is lower. This fact was reported by Selkäinaho and Collan (1984) who claim that a reliable distinction between various grades

Table 2. Classification matrix for DA10, DA11 and DA12, and coefficients of the discriminant functions.

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		CIN I	CONDYLOMA-CIN I	
CIN I	7	8	1	85.7%
Condyloma-CIN I	22	2	20	90.9%
		Total % concordance: 89.66%		

DISCRIMINANT FUNCTIONS				
	CIN I	CONDYLOMA-CIN I	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-86.11	-67.89		
NFFELL-S	375.76	333.30	11.59	0.700

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		CIN II	CONDYLOMA-CIN II	
CIN II	11	9	2	81.8%
Condyloma-CIN II	27	4	23	85.2%
		Total % concordance: 84.21%		

DISCRIMINANT FUNCTIONS				
	CIN II	CONDYLOMA-CIN II	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-729.59	-713.69		
CFFAR-S	224.57	238.24	5.73	0.658
NMaD-M	-2.03	-0.65	5.52	0.755
CFFCI-M	327.35	318.15	3.79	0.902

CYTOPATHOLOGIST'S GROUPS	No. CASES	PREDICTED GROUP		PERCENT OF CASES CORRECTLY CLASSIFIED
		CIN III	CONDYLOMA-CIN III	
CIN III	37	30	7	81.1%
Condyloma-CIN III	10	2	8	80.0%
		Total % concordance: 80.85%		

DISCRIMINANT FUNCTIONS				
	CIN III	CONDYLOMA-CIN III	EQUIVALENT F STATISTIC	WILK'S LAMBDA
Constant	-47.47	-48.72		
CArea-M	-0.60	-0.51	15.73	0.737
NArea-S	0.42	0.73	10.07	0.681
CMiD-M	14.85	13.27	8.58	0.620

NFFELL-S: nuclear form factor ELL for superficial cells; CFFAR-S: cytoplasmic form factor AR for superficial cells; NMaD-M: nuclear maximal diameter for intermediate cell; CFFCI-M: cytoplasmic form factor CI for intermediate cells; CArea-M: cytoplasmic area for intermediate cells; NArea-S: nuclear area for superficial cells; CMiD-M: cytoplasmicic minimal diameter for intermediate cells.

of pathological changes with a single morphometric measurement may be difficult due to overlap between grades. When several parameters are applied at the same time, the possibility of obtaining a clear-cut distinction between grades is more possible.

The results of the discriminant analyses in the present investigation lead to several conclusions. In general, the features chosen to better discriminate between groups were mainly cytoplasmic and, specially, nuclear size. When the discrimination process was performed on a high number of groups, the percentage of correctly classified cases decreased. The decrease in the number of diagnostic categories allows better discrimination. The quantitative microscopic classification rules increased to 100% of specificity for normal samples while the efficiency was 85-90% in correctly classifying between normal and the other three sets of pathological cases. One possible interest for applying the histo-pathological diagnosis is that only a few (1 or 2) cellular features are necessary in the discriminant functions. Thus, the quantitative statistical analysis serves to correct misinterpretations and/or to support microscopic diagnosis.

If the aim of the pathologist is to interpret what differences exist within pathological entities, other discriminant functions must be applied. We consider that a difficult separation exists when CIN grades are present. The discriminant functions establish an objective diagnosis of these processes. The same happened when condyloma-CIN grades were included in our study. In both cases, with the increase of tumoral dedifferentiation the sensitivity to correctly classify cases diminished. Most of the quantitative misclassified cases, were included in phases which were close together. For instance, from five misclassified cases of condyloma-CIN I lesions (22.7% of the total condyloma-CIN I cases), 4 were included in condyloma-CIN II group and only one in condyloma-CIN III group.

An objective of our study was to evaluate the changes in uterine cervix epithelium when it is infected by papillomavirus. The valuation of quantitative data between normal versus condyloma-NCIN groups, CIN I versus condyloma-CIN I groups, CIN II versus condyloma-CIN II groups and CIN III versus condyloma-CIN III groups makes it clear that there is an increase in both nuclear and cytoplasmic size when human papillomavirus infection occurs (Artacho-Pérula et al., 1993). Tosi et al. (1988) reported a progressive increase of nuclear area in the cells of deep, intermediate and superficial layers parallel to that of tumoral dedifferentiation; the association with human papillomavirus includes a higher rate or progression. Our findings indicate that the combination of features can be reliable in correctly classifying these entities, considering that 80-90% of the cases were identified whereas Tosi et al. (1988) could not differentiate CIN III from condyloma-CIN III using bivariate analysis (Student's t-test).

References

Abdul-Karim F.W., Fu Y.S., Reagan J.W. and Wentz W.B. (1982). Morphometric study of intraepithelial neoplasia of the uterine cervix. Obstet. Gynecol. 60, 210-214.

Artacho-Pérula E., Roldán-Villalobos R., Salas-Molina J. and Vaamonde-Lemos R. (1993). Histomorphometry of normal and abnormal cervical samples. Anal. Quant. Cytol. Histol. (15, 290-297).

Baak J.P.A., Kurver P.H.J. and Boon M.E. (1982). Computer-aided application of quantitative microscopy in diagnostic pathology. Pathol. Ann. 2, 287-306.

Baak J.P.A., Nauta J.J.P., Wisse-Brekelmans E.C.M. and Bezemer P.D. (1988). Architectural and nuclear morphometrical features together are more important prognosticators in endometrial hyperplasias than nuclear morphometrical features alone. J. Pathol. 154, 335-341.

Bibbo M., Montag A.G., Lerma-Puertas E., Dytch H.E., Leelakisolvong S. and Bartels P.H. (1989). Karyometric marker features in tissue adjacent to invasive cervical carcinomas. Anal. Quant. Cytol. Histol. 11, 281-285.

Boon M.E. and Kok L.P. (1985). Koilocytotic lesions of the cervix: the interrelation of morphometric features, the presence of papilloma-virus antigens, and the degree of koilocytosis. Histopathology 9, 751-763.

Collan Y. (1989). General principles of grading lesions in diagnostic histopathology. Pathol. Res. Pract. 185, 539-543.

Fu Y.S., Ferenzy A., Huang I. and Gelfand M.M. (1988). Digital imaging analysis of normal, hyperplastic and malignant endometrial cells in endometrial brushing samples. Anal. Quant. Cytol. Histol. 10, 139-149.

Fu Y.S., Cheng L., Huang I., Huang S., Wiesmeier E., Wettstein F. and Weissman M. (1989). DNA ploidy analysis of cervical condyloma and intraepithelial neoplasia in specimens obtained by punch biopsy. Anal. Quant. Cytol. Histol. 11, 187-195.

Gissmann L., Boshart M., Dürst M., Ikenberg H., Wagner D. and zur Hausen H. (1984). Presence of human papillomavirus in genital tumors. J. Invest. Dermatol. 83, 26-28.

Hall T.L., Philippe A. and Rosenthal D.L. (1988). Multifactor analysis of intermediate cells from the uterine cervix. The importance of slide effects on variance components. Anal. Quant. Cytol. Histol. 10, 166-170.

Hartman B., Kaplan B. and Boone D. (1986). Morphometric analysis of dysplasia in cervical cone biopsy specimens in cases with false-negative cytology. Obstet. Gynecol. 68, 832-835.

Kadish A.S., Burk R.D., Kress Y., Calderin S. and Romney S.L. (1986). Human papillomavirus of different types in precancerous lesions of the uterine cervix: histologic, immunocytochemical and ultrastructural studies. Hum. Pathol. 17, 384-392.

Marchevsky M., Gil J. and Jeanty H. (1987). Computerized interactive morphometry in pathology: current instrumentation and methods. Hum. Pathol. 18, 320-331.

Mariuzzi G.M., Montironi R., Di Loreto C. and Sisti S. (1989). Multiparametric quantitation of the progression of uterine cervix preneoplasia towards neoplasia. Pathol. Res. Pract. 185, 606-611.

Meijer C.J.L.M., van der Loo E.M., van Vloten W.A., van der Velde E.A., Scheffer E. and Cornelisse C.J. (1980). Early diagnosis of mycosis fungoides and Sezary's syndrome by morphometric analysis of lymphoid cells in the skin. Cancer 45. 2864-2871.

Montag A.G., Bartels P.H., Lerma-Puertas E., Dytch H.E., Leelakusolvong S. and Bibbo M. (1989). Karyometric marker features in tissue adjacent to in situ cervical carcinomas. Anal. Quant. Cytol. Histol. 11, 275-280.

Norris H.J., Becker R.L. and Mikel U.V. (1989). A comparative morphometric and cytophotometric study of endometrial hyperplasia, atypical hyperplasia, and endometrial carcinoma. Hum. Pathol. 20, 219-223.

Selkäinaho K. and Collan Y. (1984). Stereologic and morphometric measurements in diagnostic histopathology: statistical considerations. Acta Stereol. 3, 33-38.

Syrjänen K. (1983). Human papillomavirus (HPV) lesions in association with cervical dysplasias and neoplasias. Obstet. Gynecol. 62, 617-624.

Syrjänen K., Mäntyjärvi R., Väyrynen M., Syrjänen S., Parkkinen S., Yliskoski M., Saarikoski S. and Castrén O. (1987). Human papillomavirus (HPV) infections involved in the neoplastic process of the uterine cervix as established by prospective follow-up of 513 women for two years. Eur. J. Gynecol. Oncol. 8, 5-16.

Tosi P., Luzi P., Sforza V., Spina D., Miracco C. and Santopietro R. (1986). Morphometrical assessment of mean nuclear area in breast cancer in comparison with that of lymph node metastasis. Pathol. Res. Pract. 180, 498-501.

Tosi P., Luzi P., Santopietro R., Miracco C., Lio R., Syrjänen S., Mäntyjärvi R. and Syrjänen K. (1988). Morphometric assessment of the biological potential of human papillomavirus infections in the uterine cervix. Appl. Pathol. 6, 247-257.

Twiggs L.B., Clark B.A. and Okagaki T. (1981). Basal cell pseudopodia in cervical intraepithelial neoplasia; progressive reduction of number with severity: a morphometric quantification. Am. J. Obstet. Gynecol. 139, 640-644.